明治大学総合数理学部

2016 年度

卒　業　研　究

IP アドレスによる位置情報検索システムの開発と評価 (2) 位置情報取得の精度

学位請求者　先端メディアサイエンス学科

高 橋 俊 也

目次

第 1 章

はじめに

1.1 研究背景

昨今，情報漏洩やサイバー攻撃などの事件が頻繁に起こりサイバーセキュリティへの関心が増加している．2015 年には新潟県庁では情報漏洩には至らなかったが庁内のパソコンが「水飲み場型」攻撃を受け海外のサイトへ自動的にアクセスしてしまう [1] など，被害や危険性は広がっている．2016 年には神奈川や京都でも不正サイトによる被害があった．このように日本全国でサイバー攻撃の被害が発生している．よって私たちは都道府県ごとのサイバー攻撃の危険度を示し，セキュリティ意識の向上に役立てる．

研究当初は前述した問題に対し危険度の指標を作り都道府県ごとの危険度を可視化することが研究目標だったが，指標の定義やサイバー攻撃に関するパケット収集の難易度の高さから研究目標をパケットの IP アドレスについて調査した際に使用した GeoIP ロケーションサービスに変更した．IP アドレスから，その IP を用いる Web サーバの位置を提供する GeoIPTool[2] などの GeoIP ロケーションサービスがある．これらのサービスは Web サイトの管理者がどのようなプロバイダを使用しているユーザがサイトにアクセスしているのかを調べたり，不正な通信がどの国から送られてきているのかを調べるなど様々な使い道がある．だが MaxMind などの GeoIP ロケーションサービスはデータベースの精度が悪く登録されている IP アドレスも少ない．よって GeoIP ロケーションサービスで有用な情報を得ることは難しい．

そこで，GeoIP サイトを利用しているユーザがその時使用している IP アドレスとその位置情報を登録できる GeoIP ロケーションシステムを開発した．自分の位置情報を登録するためには幾つかのキーワードで google マップを用いて手動で現在地を探さなくてはならない．そこで，本研究ではこの問題を解決するために現在地を自動的に取得する機能を実装した．本稿では，この位置情報取得の評価，またそれらを使って集めたデータの分析結果について述べる．

1.2 研究目的

本論文では位置情報取得の精度を分散分析を用いて分析することにより測定することと，その精度が登録された情報にどのような影響を与えるのか分析することを目的とする．

1.3 先行研究"次元圧縮によるダークネットトラフィックデータの可視化"

不正な通信の収集と分析の手段として北園淳らはダークネットトラフィックと t-SNE を提案している [3].

1.3.1 研究背景

特定のホストが割り当てられていない IP アドレス空間はダークネットと呼ばれる．ダークネットには本来，パケットが到達することはないはずだが実際には送信元を偽装した DDoS 攻撃に対する返信や，マルウェアによるスキャンなど，不正な活動に使われるパケットが大量に届く．ダークネット空間に流れてくるパケットをダークネットトラフィックと呼ぶ．

ダークネットトラフィックは，その元になる不正な活動の多様性に反映して，複雑で多様なパターンをとる．そのパターンは高次元ベクトルとして表現できるがダークネットトラフィックは膨大な量のため，時々刻々と変化するパターンの分布や傾向を把握するのは困難である．

1.3.2 提案手法

ダークネットトラフィックパターンの分布や傾向を，一見して直感的に捉える事を目的として t 分布型確率的近傍埋め込み法 (t-Distributed Stochastic Neighbor Embedding, t-SNE) と呼ばれる次元圧縮手法をダークネットトラフィックに対して適用して，ホスト毎のトラフィックパターンを可視化する．

1.3.3 t-SNE

次元圧縮手法 t-SNE は高次元データの任意の 1 行とその他の行の類似度を正規分布や t 分布の確率分布を利用し，低次元のグラフの点同士の距離として表す．t-SNE は，元の高次元データに内在するデータ分布の構造自体も高次元の場合に特に有効である．

1.3.4 t-SNE によるダークネットトラフィックの可視化

情報通信研究機構によって提供されている MWS データセットを用いた．このデータセットのうち 2014 年 1 月 1 日から 2014 年 4 月 30 日にかけて，4 カ月に渡って収集されたダークネットパケットを以下の 17 種類の特徴を抽出して用いる．ダークネットパケット DDoS 攻撃のバックスキャッタもしくはスキャンが多くを占めるが本研究のシステムでは，2 つの判別が難しいのでそれ以外のパケットに注目し可視化する．

- 総パケット数
- 用いられているプロトコルの種類の数
- ペイロードに関するもの
  - ペイロードサイズの平均
  - ペイロードサイズの分散
- 時間に関するもの
  - パケット間の時間間隔の平均
  - パケット間の時間間隔の分散
- 宛先に関するもの
  - 宛先 IP アドレスの総数
  - 宛先 IP アドレスごとのパケット数の平均
  - 宛先 IP アドレスごとのパケット数の分散

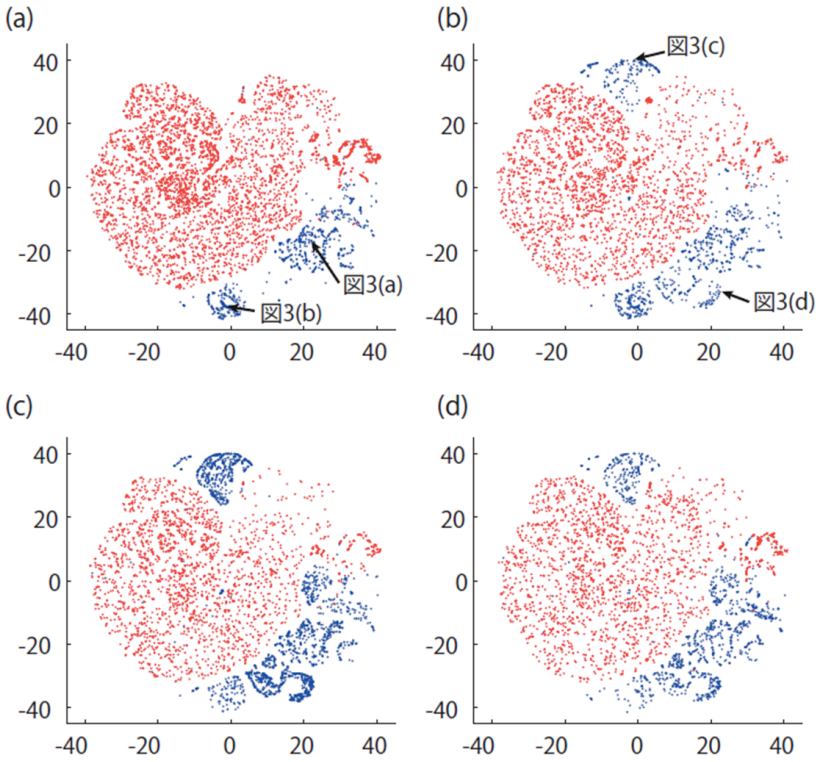

図 1.1: t-SNE による可視化の結果

– 連続するパケット間の宛先 IP アドレスの差分の平均
– 連続するパケット間の宛先 IP アドレスの差分の分散
– 宛先ポートの総数
– 宛先ポートごとのパケット数の平均
– 宛先ポートごとのパケット数の分散

上記の特徴量のうち，総パケット数，パケット数の平均・分散については，値が大きくなるため自然対数を取る．また，値を [0, 1] の範囲に規格化する．

図 1.1 に t-SNE を適応した結果を散布図として示す．図 (a)-(d) はそれぞれ 1-4 月に分けてプロットしたものである．赤色の点は DDoS 攻撃のバックスキャッタ，青色はそれ以外を表す．いずれの月においても 2 色の点が広がる領域が綺麗に分かれており，用いた 17 種類の特徴はトラフィックの違いを捉えることができていると言える．

1.3.5 まとめ

本研究では，ダークネットトラフィックから特徴を取り出し，t-SNE を用いることでパターンの分布を散布図として可視化した．本研究で行った可視化を用いることによって，2008 年に起こった Conficker の流行のような，新たな不正通信パターンの発生や急増を早期に捉えることが期待される．

第 2 章

Prefecture maP Internet Protocol

2.1 システム概要

PPIP はユーザが位置情報や IP アドレスを登録でき，情報を検索することができる下記の二つの機能を持ったシステムである．

機能 1. Geolocation API を利用し，ユーザの現在地を取得し，PPIP データベースに登録する．
機能 2. PPIP データベースから登録されているデータを検索する．

PPIP システム開発担当
高橋　Google Maps JavaScript API[6],Geolocation API[7],TLS 公開鍵証明書の導入
笹　　Google Maps JavaScript API[6],Google Places API[8]
厚見　データベースの設計・作成 [9]，データのやりとり [9]

2.2 使用 API

2.2.1 Geolocation API

Geolocation API はデバイスの位置情報を携帯電話の基地局や Wi-Fi のアクセスポイント，GPS(Global Positioning System) などを利用し，緯度経度の値として取得する API である．GPS を搭載していないパソコンやタブレットなどでも Wi-Fi のアクセスポイントなどを利用して位置情報を取得する．セキュリティの観点より，位置情報を取得する場合は TLS 公開鍵証明書により認証されたサーバからの暗号化された通信でなければならない．

GPS による位置情報取得

GPS は携帯電話，スマートフォンなどに搭載されている自分の現在位置を測定するためのシステムである．図 2.1-2.3 に位置情報測定の仕組みを示す．GPS を搭載したデバイスが宇宙にある衛星 A から衛星 A の位置とメッセージ発信時刻の情報を含む電波を受信し，それらの情報とメッセージ受信時刻，電波の速度を用いてデバイスと衛生 A 間の距離 A を測定する．デバイスは衛星 A を中心とした半径 A の円球上にある．衛星 B と C についても求め円球が重なるところにデバイスの位置情報を測定する．

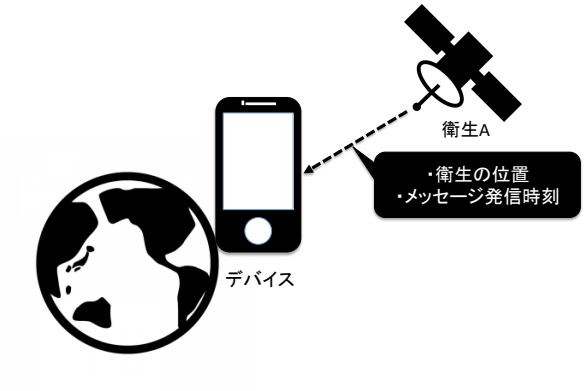

図 2.1: GPS による位置情報取得の仕組み

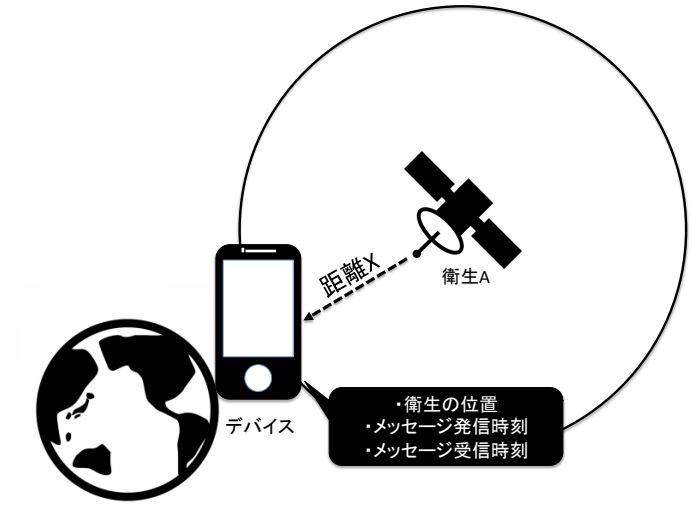

図 2.2: GPS による位置情報取得の仕組み

Wi-Fi のアクセスポイントによる位置情報取得

衛星からの電波が届きにくい屋内で位置情報を取得する際や，GPS が搭載されていない PC やタブレットで位置情報を取得する際には Wi-Fi のアクセスポイントを用いる．図 2.4-2.5 に位置情報測定の仕組みを示す．デバイスの周囲にある Wi-Fi のアクセスポイントから発信される電波を受信し，その電波の強度，そして各デバイスを提供する会社が管理するデータベースから分かるそのアクセスポイントの位置情報を用いてデバイスとアクセスポイント間の距離を取得する．周囲に多くの Wi-Fi のアクセスポイントがあれば精度は高くなる．

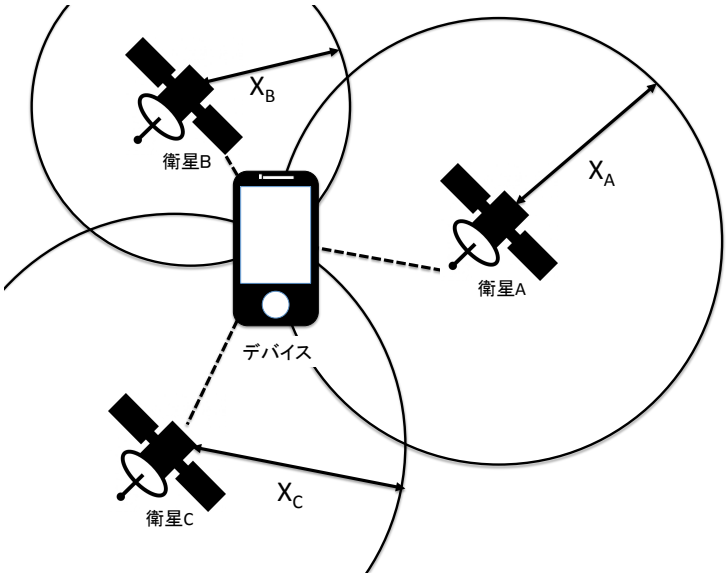

図 2.3: GPS による位置情報取得の仕組み

図 2.4: Wi-Fi のアクセスポイントによる位置情報取得の仕組み

携帯電話の基地局による位置情報取得

Wi-Fi のアクセスポイントによる位置情報取得と同様に，周辺の基地局の電波を受信し，電波の強さと基地局の位置情報からデバイスと基地局間の距離を測定して位置情報を取得する．

2.2.2 Google Maps JavaScript API

IP アドレスによる位置情報システムの開発と評価 (1)，2.2.2 を参照．

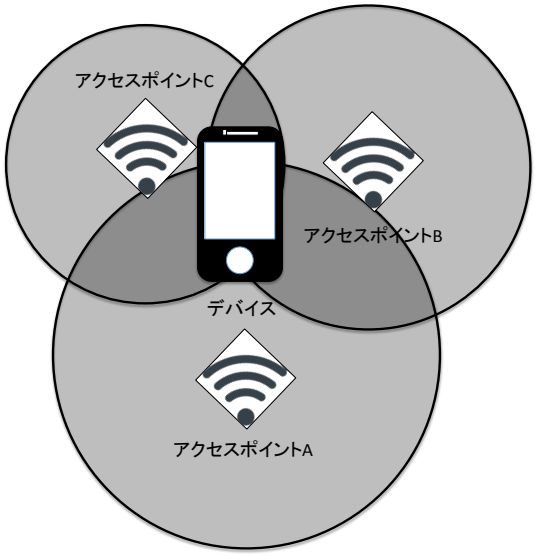

図 2.5: Wi-Fi のアクセスポイントによる位置情報取得の仕組み

2.2.3 Google Places API

IP アドレスによる位置情報システムの開発と評価 (1)，2.2.3 を参照.

2.3 登録システム

IP アドレスによる位置情報システムの開発と評価 (1)，2.3 を参照.

2.4 検索システム

IP アドレスによる位置情報システムの開発と評価 (1)，2.4 を参照.

第 3 章

TLS 公開鍵証明書

3.1 現在地取得と暗号化通信

今日，ユーザの位置情報はプライバシー情報の 1 つとして盗聴などの攻撃の対象になっている．Google map API ではプライバシー情報の漏洩を未然に防ぐ為，暗号化されていない通信上での位置情報取得は禁止されている．PPIP では TLS 公開鍵証明書を用いた暗号化通信を行っている．本章では TLS 公開鍵証明書による通信の仕組みと証明書の発行と設定について述べる．

3.2 TLS とは

TLS(Transport Layer Security) とは、インターネット上で通信を暗号化する技術である．SSL(Secure Sockets Layer) を元に作られたもので SSL/TLS と呼ばれる場合もある．通信を暗号化することでデータの盗聴や改ざんを防ぐことができる．この通信を実現するには認証局でサーバ証明書を発行し設定する必要がある．

3.3 TLS の仕組み

TLS による暗号化された通信の仕組みを図 3.1 に示す．Web サイトを管理するサーバ側は秘密鍵と公開鍵を持ち認証局による証明書を持つとする．Web を閲覧するクライアント側はサーバ側に接続要求を送る．要求を受け取ったサーバ側はクライアントにサーバ証明書と公開鍵を送る．クライアント側は認証局が発行するルート証明書を用いてサーバ証明書が正当なものか検証する．クライアント側は検証が成功するとサーバとの通信に用いるための共通鍵を生成する．これをサーバ証明書と一緒に送られてきた公開鍵を用いて暗号化し，サーバ側に送る．サーバ側は受け取った共通鍵を秘密鍵を用いて復号する．これで両者は共通鍵を持つ．この共通鍵を用いて暗号化通信を行うことでデータの盗聴や改ざんを防ぐことができる．

3.4 CSR の作成

CSR(Certificate Signing Request) は認証局に証明証を発行してもらうためにサーバの所在地等の情報を含んだデータである．これを作成して認証局に申請することで証明書を発行する．CSR の作成方法を示す．Apache2.x と mod ssl，OpenSSL を使用し，認証局「GlobalSign」を利用した際の作成方法である．また今

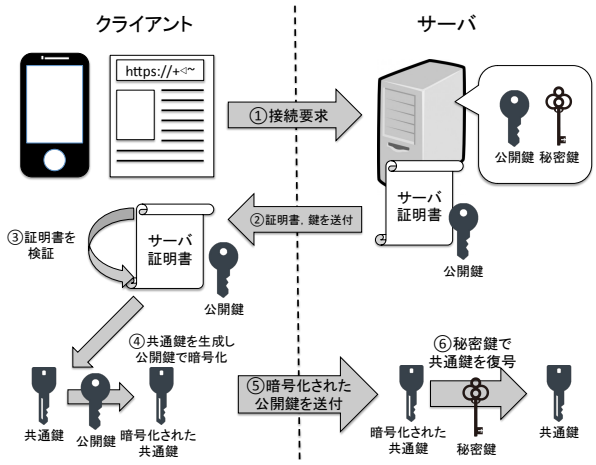

図 3.1: TLS による暗号化された通信

表 3.1: CSR 作成環境

コモンネーム	ssl.ppip.com
conf ディレクトリまでのパス	/etc/httpd/conf/
秘密鍵の保存ディレクトリ	/etc/httpd/conf/ssl.key/
CSR の保存ディレクトリ	/etc/httpd/conf/ssl.csr/
秘密鍵のファイル名	ssl.ppip.com.key
CSR のファイル名	ssl.ppip.com.csr

回の作成方法ではサーバによって異なるコモンネームなどは表 3.1 とする．

Apache の conf のパス上で，プログラム 3.1 を実行し，秘密鍵を生成する．次にプログラム 3.2 を実行し，プログラム 3.3 に必要情報を入力し，CSR を作成する．入力項目を表 3.2 に示す．CSR を認証局に送付し証明書の発行を待つ．

プログラム 3.1: 秘密鍵の生成 [10]

```
openssl genrsa -des3 -out ./ssl.key/ssl.ppip.com.key 2048
```

プログラム 3.2: CSR の生成 [10]

```
openssl req -new -key ./ssl.key/ssl.ppip.com.key
 -out ./ssl.csr/ssl.ppip.com.csr
```

プログラム 3.3: 情報の入力 [10]

```
Country Name (2 letter code) [GB]:
State or Province Name (full name) [Berkshire]:
Locality Name (eg, city) [newbury]:
Organization Name (eg, city) [My company Ltd]:
Organization Unit Name (eg, section) []:
Common Name (eg, your name or your server's hostname) []:
Email Address []:
```

表 3.2: 入力項目と入力例 [10]

入力項目	内容	入力例
Country Name	国を示す 2 文字の ISO 略語	JP
State or Province Name	組織が置かれている都道府県	Tokyo
Locality Name	組織が置かれている市区町村	Nakano-ku
Organization Name	組織の名称	Meiji University
Organization Unit Name	組織での部署名	Kikuchi Lab
Common Name	ウェブサーバの FQDN	ssl.ppip.com
Email Address	入力不要	-
A challenge password	入力不要	-
An optional company name	入力不要	-

表 3.3: 証明書インストール環境

コモンネーム	ssl.ppip.com
conf ディレクトリまでのパス	/etc/httpd/conf/
SSL 設定用 conf ファイル	/etc/httpd/conf.d/ssl.conf
証明書の保存ディレクトリ	/etc/httpd/conf/ssl.crt/
サーバ証明書ファイル名	ssl.ppip.com.crt
中間 CA 証明書ファイル名	dvcacert.cer

```
please enter the the following 'extra' attributes
to be sent with your certificate request
A challenge password []:
An optional company name []:
[root‘ssl conf]#
```

3.5 TLS 公開鍵証明書の設定

認証局より証明書を発行したらサイトにインストールすることで通信が暗号化される．CSR の作成と同様に Apache2.x と mod ssl，OpenSSL を使用した際の作成方法である．インストール環境は表 3.3 に示す．証明書はサーバの管理者に送られるので証明書，中間 CA 証明書のデータをサーバに保存する．プログラム 3.4 を実行し，証明書をインストールする．Apache のデーモンを再起動すると暗号化通信が可能になる．

プログラム 3.4: ssl の設定 [10]

```
SSLEngine on
SSLCertificateChainFile /etc/httpd/conf/ssl.crt/dvcacert.cer
SSLCertificateFile /etc/httpd/conf/ssl.crt/ssl.ppip.com.crt
SSLCertificateKeyFile /etc/httpd/conf/ssl.key/ssl.ppip.com.key
```

第 4 章

Free Wi-Fi のセキュリティ調査

4.1 実験概要

当初の研究計画において危険度の指標を決めるため，2016 年 8 月に東京都新宿駅周辺，中野駅周辺，埼玉県大宮駅周辺，鴻巣駅周辺でそれぞれ 10〜15 個の Free Wi-Fi のセキュリティ調査を行った．本実験は我々も使用する可能性がある Free Wi-Fi のセキュリティの有無，流れているパケットから Free Wi-Fi の危険性を明らかにすることを目的とする．

4.2 Free Wi-Fi とセキュリティ

Free Wi-Fi とは会員登録やゲストコードの入力などにより誰でも無料でインターネットにつなぐことができる Wi-Fi のことである．カフェやコンビニ、空港、ファストフード店，商業施設など様々な場所に設置されている．セキュリティが設定されていない場合は通信内容は暗号化されず，通信を傍受された場合には個人情報が流失してしまうなどの危険がある．ただしサイトなどを閲覧する際に Web サーバが TLS 公開鍵証明書により認証されている場合などにい関しては通信が暗号化される．接続の容易を優先した結果セキュリティを設定しない Free Wi-Fi が多いのではないかと考えられる．

4.3 実験方法

Free Wi-Fi に接続しデバイスのネットワーク環境を確認する画面からセキュリティの有無を調べる．加えて Wireshark で 10 分間パケットを観測する．

4.4 実験結果

Free Wi-Fi のセキュリティの有無についての実験結果を表 4.1 に示す．今回調査した Free Wi-Fi でセキュリティの設定がされているのは全体の 5 パーセントであった．これよりほとんどの場合 Free Wi-Fi にはセキュリティが設定されていないことが分かる．セキュリティが設定された Wi-Fi を使用する場合，パスワードを入力する必要がある．今回の調査でセキュリティが設定されていた 2 つの Free Wi-Fi のパスワードは Free Wi-Fi が設置されている店内や Web サイトにパスワードが書かれていたが日本語に不自由な観光客や Wi-Fi に詳しくない人は Wi-Fi 接続時にパスワードを見つけられず諦めてしまう場合もあると考えられる．

表 4.1: 実験結果

場所	調査数	セキュリティあり	セキュリティなし
新宿	15	1	14
中野	10	0	10
大宮・鴻巣	10	1	9

第 5 章

データ収集実験

5.1 実験概要

IP アドレスと位置情報の関係を明らかにするために，PPIP を公開しデータ収集を行った．SNS で公開したので実験協力者は主に大学生である．

5.2 実験結果

明治大学菊池研究室内で行った予備実験と 2016 年 12 月，オープンな環境での本実験を行ったところ 138 個の IP アドレスとその位置情報が収集された．

5.2.1 IP アドレスの種類による分類

登録された IP アドレスを種類別に表 5.1 に示す．1 番多いのは市町村や駅の名前で登録された IP である．これは多くの実験協力者が自宅の Wi-Fi につないだ状態で実験に協力してくれたためだと思われる．次いで飲食店，商業施設，コンビニの IP アドレスの順であった．これらより普段ユーザがどのような場所で Wi-Fi を使っているかがわかる．

5.2.2 IP アドレスのプロバイダによる分類

表 5.2 に IP アドレス数を都道府県別に，図 5.1 に IP アドレス数をプロバイダ別に，図 5.2 に IP アドレスを Whois[2] を用いてプロバイダ別に分類して日本地図上に表示する．今回の実験では東京都，特に新宿，中野で登録された IP アドレス数が 78 個で一番多いので東京都にプロバイダが集中している．

表 5.1: IP アドレスの種類による分類

市町村・駅	飲食店	商業施設	コンビニ	その他
57	22	21	17	30

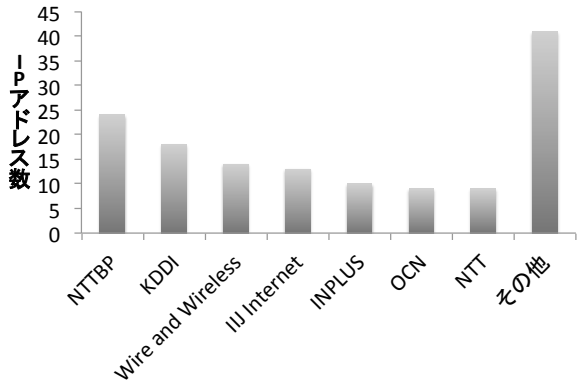

図 5.1: IP アドレスのプロバイダによる分類

表 5.2: IP アドレスの都道府県による分類

東京都	神奈川県	埼玉県	その他
78	27	18	15

図 5.2: プロバイダーを地図上にプロット

第 6 章

位置情報取得機能実験

6.1 実験概要

PPIP で位置情報を取得する際にどれほどの誤差が生じるのか，またそれにより登録情報にどのような影響を及ぼすのか明らかにすることを目的とする．Iphone5s(GPS 機能あり)，iPad(GPS 機能なし)，Macbook air(GPS 機能なし) の 3 つのデバイスを用いて PPIP を開いて現在地を取得する．その結果について述べる．

2016 年 12 月 1 日，2 日に 3 つのデバイスを用いて明治大学中野キャンパスの研究室と埼玉県の自宅で実際の位置情報とどれほどずれがあるのかそれぞれ 10 回ずつ実験を行った．PPIP で手動で登録した際の緯度経度を正しい位置情報とする．また Geolocation API に GPS の利用や電力消費量の制限をなくした 2 種類の精度があるのでそれについても比較する．

6.2 実験結果

位置情報取得の誤差についての結果を表 6.1 に，精度についての結果を表 reftable.result2 に示す．精度の信頼度は 90 % とする．誤差については iPhone5s が一番小さい．これは 3 つのデバイスの内，唯一 GPS を内蔵しているからだと思われる．分散についても 12 種類の環境下の内 8 種類は 40 以下の値となっており，測定結果が安定していることがわかる．

位置情報取得時のデバイスの違いによる影響を分析するために分散分析を行った．その結果，研究室で位置情報取得した場合はデバイスの違いの効果は有意であった ($F(2,57)=4.81$, $p < .05$)．多重比較により，「PC」と「iPhone5s」，「PC」と「iPad mini」の間に有意差があり，iPhone と iPad mini は PC と比較して誤差が小さい事がわかった．

位置情報取得時の場所の違いによる影響を分析するために分散分析を行った．その結果，場所の効果は有意で ($F(1,118)=468.77$, $p < .01$)，自宅で位置情報取得した時の誤差が研究室での誤差より小さい事がわかった．

精度については全体では実験全体では 60m，最も小さい精度は自宅で実験した場合の 19m であった．デバイスで比べると誤差と同様に iPhone5s が 1 番小さい．プログラム内の精度の設定については差は見られなかった．これらのことより精度はプログラム内の設定よりもデバイスや位置情報取得する場所の要因が大きいのことが考えられる．

最大で 100m 弱の誤差が生じたが登録情報には位置情報に加えて，建物や地域の名前も含まれているのでユーザはそれらの情報と合わせて IP アドレスの位置情報を読み取ることができる．よって登録情報に及ぼす影響は小さいと考える

表 6.1: 実験結果：誤差

場所	デバイス	精度	誤差 [m]	最大誤差	分散
研究室	PC	低	82.7	85.8	4.0
		高	83.5	85.7	1.9
	iPhone5s	低	59.7	66.4	14.4
		高	58.5	67.5	38.5
	iPad mini	低	63.7	91.2	336.7
		高	60.0	75.8	282.4
自宅	PC	低	15.2	17.2	1.3
		高	15.1	17.0	0.6
	iPhone5s	低	23.5	50.6	76.4
		高	14.5	24.4	34.2
	iPad mini	低	20.7	22.7	1.9
		高	18.0	21.5	2.8

表 6.2: 実験結果：精度

実験環境	精度
精度：高	± 60m
精度：低	± 60m
場所：研究室	± 61m
場所：自宅	± 19m
デバイス：PC	± 60.5m
デバイス：iPhone5s	± 53m
デバイス：iPadmini	± 61m
全体	± 60m

第 7 章

おわりに

既存の GeoIP サービスは IP アドレスの登録数が少なく精度が悪いため，ユーザがその時使用している IP アドレスとその位置情報を登録できる GeoIP システム「PPIP」を開発した．位置情報の自動取得機能を追加し，その精度を求めた．結果として全体の精度は信頼度 90 %で 60m であった．また精度の大きな要因はデバイスの違いや位置情報取得する際の場所の環境であると思われる．

本実験で集めたデータからは IP アドレスと位置情報の関係の特徴は発見できなかったが，より多くの IP アドレスの情報が登録されれば特定の地域に多いプロバイダや，ある店の Wi-Fi に多く使われているプロバイダ，例えばマクドナルドやロッテリアなどのファストフード店の Wi-Fi は全て Softbank がプロバイダである，といった特徴がわかるだろう．

参考文献

[1] ITpro，新潟県庁にサイバー攻撃，(http://itpro.nikkeibp.co.jp/atcl/news/15/061802053/?rt=nocnt，2016 年 4 月参照).

[2] Geo IP Tool(https://geoiptool.com/，2016 年 10 月参照).

[3] 北園淳，古谷暢章，宇川雄樹，班涛，中里純二，島村隼平，小澤誠一，“次元圧縮によるダークネットトラフィックデータの可視化”，SCIS2016，2016 年 4 月参照

[4] 竹下-恵，パケットキャプチャ入門-第 3 版-LAN アナライザ Wire-shark 活用術，2016 年 6 月参照

[5] Whois(http://www.cman.jp/network/support/ip.html，2016 年 10 月参照).

[6] Syncer (https://syncer.jp/google-maps-javascript-api-matome 2016 年 10 月参照)

[7] Geolocation(http://www.htmq.com/geolocation/，2016 年 10 月参照).

[8] VINTAGE (http://www.vintage.ne.jp/blog/2015/04/395 2016 年 10 月参照)

[9] PHP(http://php.net/manual/ja/langref.php　2016 年 10 月参照)

[10] GlobalSign，サポート：SSL サーバ証明書 (https://jp.globalsign.com/support/ssl/)，2016 年 12 月参照

[11] PPIP(http://windy.mind.meiji.ac.jp/ ksa/senior/top.php 2016 年 10 月開発)

第 8 章

謝辞

本研究に際し，熱心にご指導いただいた指導教官の菊池浩明教授，実験に協力してくださった研究室のみなさんに感謝申し上げます．

付録 A

3 年次の研究”会員カードユーザの「プライバシ侵害」に対する意識の研究”

A.1 研究背景

「IC カードを使って電車に乗り，コンビニで買い物をしてポイントカードにポイントを貯める」といったように，我々の生活において IC カード，ポイントカードなどの会員カードを使用する機会が増加している．それらの会員カードを使用する際には，会員の購買行動などの様々な履歴がカード事業者に取得されている．

2013 年 7 月，JR 東日本は自社が運営する Suica 4300 万枚の乗降履歴の販売を開始し，大手家電メーカーが購入することが発表された．個人情報の漏洩を恐れたユーザ約 5 万人がデータを使用しないように申し入れを行う「オプトアウト」を行った．このように会員カード事業者と顧客の間に提供してもいい情報の違い，プライバシ侵害に対する意識の違いが存在している．

A.2 研究目的

本研究は [1] において，小松らが行ったアンケートの結果と，今年 11 月に同一被験者に行ったアンケート結果の比較から会員カードユーザの意識の変化を研究することを目的とする．ユーザが履歴販売についての理解がどのように変化しているのかを明らかにするために「2013 年に JR が行った Suica 利用履歴販売において販売された情報にどのようなものが含まれると思うか」を選択方式で新たに追加した．

A.3 アンケートによる実験

2015 年 11 月に明治大学総合数理学部の学生 10 名にアンケートを実施した．アンケートは 5 項目に分かれている．

A.4 実験結果

A.4.1 会員カード使用状況と会員カードに対する意識

アンケート結果を表 A.1 に示す．参加者全員が Suica,PASUMO といった IC カードを持っていて，それらについて 7 割以上の参加者が安心感，信頼感を感じ，生活に欠かせないと回答している．

	SUICA	PASMO	WAWON	T-POINT	PONTA	NANACO	EDY	当てはまるものはない
使用したことがある	8	2	0	7	2	1	0	0
生活に欠かせない	7	2	0	2	0	0	0	0
安心感を感じる	6	1	0	1	0	0	0	3
信頼感を感じる	7	1	0	2	0	0	0	2
実は使いたくない	0	0	0	1	0	0	0	10
実は使いたい	0	0	0	1	1	3	3	3
好感を持っている	6	1	1	3	1	1	0	2

表 A.1：会員カード使用状況と会員カードに対する意識についてのアンケート結果

平均金額(円)			絶対にイヤと回答した人数(人)		
	2014年1月(125名)	2015年11月(10名)		2014年1月(125名)	2015年11月(10名)
携帯電話	44083	1100	クレジットカード	115	7
閲覧履歴	40873	1100	銀行口座	113	7
テスト成績	38387	17600	住所録	99	6
住所録	37520	1500	時系列位置関係	85	6
位置	36163	2200	位置	85	6
現住所	34622	2600	写メ	82	7
時系列位置	34024	2200	時系列位置	82	6
時系列位置関係	33450	2200	現住所	80	2
最寄り駅	32689	1000	発言履歴	76	4
通話履歴	30894	5650	通話履歴	73	5

表 A.2：属性に対する見積もり金額

A.4.2 個人情報提供に対する見積もり金額

個人情報提供に対する見積もり金額についてのアンケート結果上位 10 個を表 A.2 に，全体の比較を図 A.1 に示す．アンケートの実施人数に少し差はあるが，個人情報を絶対に提供しないと回答した人数は，2014 年と同一である．最頻回答項目はクレジットカードと銀行口座である．約２年が経っても，提供したくない個人情報は特に変化がないこと，また個人情報を提供する際の金額については，図 1 からも明らかなように，趣味を覗いて最大で在学校が 0.9 パーセントまで金額が下がっている．

A.4.3 プライバシ指向性尺度

プライバシ指向性尺度 [2] についてのアンケート結果を表 A.3 に示す．独居 (一人の時間を大切にする指向性) に特に高い点数が，隔離 (社会から隔絶して生活したい指向性) に低い点数が見られた．よって本アンケートに参加したユーザは一人の時間を大切にするが，人里離れて暮らしたいなどの指向性はないことが分る．また，他の指向性と合わせても，2014 年と大きな変化は見られない．

個人情報を提供するかどうかの違いがプライバシ指向性尺度の得点に与える影響を比較するため，指向性尺度と個人情報の見積もり金額の分散分析を行った．結果を表 A.4 に示す．独立変数は個人情報に対する対価の違い「提供する（報酬あり)」「提供する（報酬なし)」「提供しない」の 3 水準，従属変数はプライバシ指向性尺度の点数とする．

表 3 の網掛け部分が確率検定において有意差が見られた項目である．例えば，友人親密（なんでも話せる親しい友人を持つことは大切であるという指向性）と家族構成については，家族構成を無料もしくは有料で提供するユーザは絶対に提供したくないと回答したユーザより友人親密の得点が有意に高いということがわかる．

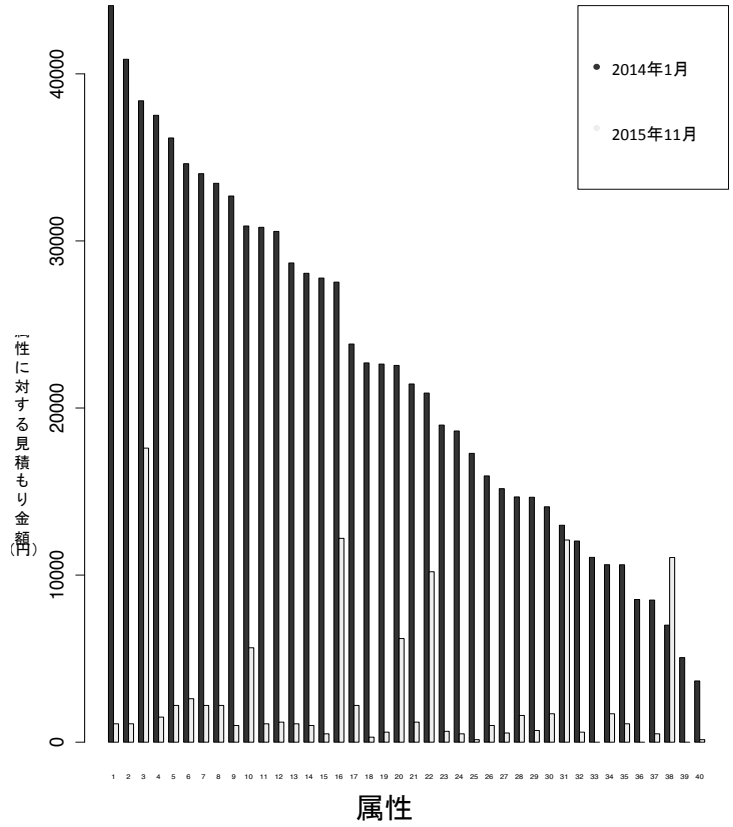

図 A.1：属性に対する見積もり金額

	独居	自由意志	友人親密	遠慮期待	家族親密	閑居	隔離
2015年11月平均点	19	15	12	12	12	13	9
2014年1月平均点	17	16	16	14	13	13	10

表 A.3：プライバシ指向性尺度平均点

A.4.4 架空のサービスに対する印象

架空のサービスに対する印象についてのアンケートでは，架空のサービス 9 項目それぞれに対して 7 件法で回答してもらった．結果を平均点で表 A.5 に示す．一番ユーザが使いたいと思っているアプリは嫌いながいる警告通知サービスであった．また全体の平均点については，2014 年の結果と比べて約 2．5 点下がっている．

架空のサービスを使いたいと思っているかどうかの違いがプライバシ指向性尺度の得点に与える影響を比較するため，架空のサービスの印象とプライバシ指向性尺度の分散分析を行った．その結果を表 A.6 に示す．独立変数は個人情報に対する対価の違い「アプリを使いたい（5 7 点)」,「どちらでもない（4 点)」,「アプリを使いたくない（1 3 点)」の 3 水準，従属変数はプライバシ指向性尺度の点数とする．

表 A.6 において，網掛け部分が確率検定において有意差が見られた項目である．例えば友人親密の得点が高

	独居	自由意志	友人親密	遠慮期待	家族親密	閑居	隔離
顔写真				3.916			
年齢							
家族構成			13.09**			4.252	
体重				3.674			
血液型	-	-	-	-	-	-	-
誕生日			8.414*				
趣味				4.59			
出身校			6.4*	5.851*		3.649	
在学校名							
使用路線		4.391					
健康状態							
彼女/氏履歴			10.17**				
彼女/氏歴							
最寄り駅		3.677		5.493*			
現住所							
携帯電話番号				6.407*			
メールアドレス				4.655			
LINEアカウント					7.38*		
skypeアカウント					3.512		
SNSアカウント							
銀行口座番号							
クレジットカード				3.571		4.388	5.011
行きつけ							
勤務先					6.322*		
発言リスト			6.177*	8.544*			
購入履歴				3.346			
閲覧履歴	5.787*					3.399	
ネット購入履歴		4.629		3.466			3.28
乗降履歴				3.636	11.36**		
位置				3.645	5.265*		
通話履歴				6.166*			
住所録				4.59			
写メ							
成績							
アプリ履歴	8.33*						
受験校名	3.793						
犯罪歴		4.864					
家族犯罪歴	***						

表 A.4：指向性尺度と個人情報に対する見積もり金額についての分散分析結果

	誕生日に応じてクーポン	購入に応じて健康アドバイス	DNA病気予測＝メニュー	学生に対するクーポン	嫌な人がいる警告	動画観覧履歴ベース推薦	写真で恋愛判定	メール内容からの推薦	SNS投稿からの推薦
2015年11月平均点	4.2	3.9	4.1	4.8	5.1	3	4.8	3.5	3.4
2014年1月平均点	5.2	3.9	4.4	5.5	4.9	3.7	4	3.8	3.6

表 A.5：架空のサービスに対する印象についてのアンケート結果

いユーザは，「自分が在学中の学校の学生だけに対して割引される情報を通知してくれるサービス」を使いたいと思っている割合が高い．だがその中でも，「在学中の大学名」を提供したくないと回答しているユーザは，そのサービスを利用したいと思っていない．

また，嫌いな人がいる警告通知サービスはほとんどのユーザが使いたいと回答したが，自分と嫌いな人との位置を図るための位置情報や元彼氏元彼女など嫌いな人になりうる人物の情報については，絶対に提供しないと回答したユーザが多い．つまり，これらのユーザは自分の位置情報は提供したくないが，嫌いながいる警告通知サービスは受けたいと考えている．このような矛盾が起きてしまうのは，ユーザがそのサービスを受けるために自分がサービスを供給する事業者に対して何の情報を提供しなければならないかを理解していないからだと思われる．このような状況が JR の履歴販売時に起きたオプトアウトの一つの原因にもなりうる．

	独居	自由意志	友人親密	遠慮期待	家族親密	閉居	隔離
誕生日にクーポン		4.375			5.567*		
購入に応じて健康アドバイス							
DNA病気予測							
学生に対するクーポン			7.254*				
イヤな人がいる警告							-
動画閲覧履歴ベース推薦							
写真での恋愛判定							
メールの内容からの推薦	9.873**						
SNS投稿からの推薦	6.45*						

表 A.6：架空のサービスの印象と個人情報に対する金額についての分散分析結果

年齢	性別	住所	血液型	乗降駅	乗降日時	家族構成
8	8	1	0	9	8	0
自宅の電話番号	携帯電話番号	顔写真	最寄り駅	SuicaのID	定期券の範囲	コンビニ
1	1	0	2	4	6	3

表 A.7：ユーザの履歴販売についての予想アンケートの結果

A.4.5 ユーザの履歴販売についての予想

ユーザの履歴販売についての予想結果を表 A.7 に示す．実際に履歴販売の対象となった項目は「年齢」，「性別」，「乗降した駅」，「乗降した日時」である．多くのユーザがこれらに対して提供されたことがわかっていたが，この他に，「定期券の範囲」や「コンビニでの Suica による買い物情報」なども販売されていると思っているユーザがいることがわかった．このようなにユーザに伝わっていない部分があることが，企業とユーザの間に意識の違いを生むのかもしれない．

A.5 おわりに

本実験では，アンケートの実施と比較分析した実験結果より，約２年前に実施した結果と比べてあまり変化が確認されなかった点，変化がされた点があった．情報提供に関する金額や使いたいアプリなどには変化が見られ，絶対に提供したくない個人情報はあまり変化が見られなかった．それらの関係を観察すると情報提供の有無と使いたいアプリの間に矛盾が確認される．これは約２年前に実施したアンケートの結果でも確認されているので，やはりユーザと企業の間の意識の違いは現在も存在しているということがわかる．

参考文献

[1] 小松孝徳, 菊池浩明，”会員カードユーザが『プライバシ侵害』を感じる理由とは？”，第 28 回人工知能学会，2014，2015 年 4 月参照

[2] 吉田圭吾，溝上慎一，”プライバシー志向性尺度 (本邦版) に関する検討”，心理学研究，2013，2015 年 4 月参照

[3] Animal Breeding Genetics 分散分析 (http://www.agri.tohoku.ac.jp/iden/toukei7.html)，2015 年 4 月参照

[4] 豊田秀樹，データマイニング入門,Pp.209-240,2015 年 5 月参照