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とシステムの開発と評価
(3) 動的 IP アドレスの調査

学位請求者　先端メディアサイエンス学科

厚 見 隆 之 介

目次

第 1 章

はじめに

1.1 研究背景

IP アドレスからその場所を提供する IP Geolocation サービスが多く存在する．たとえば GeoIP Tool[1] や GeoIP MaxMind[2] など IP アドレスを入力し，アドレスの現在地を提供している．しかし，提供されるデータベースには十分な IP アドレスの登録がなく，その精度も低い．さらに契約しているプロバイダーによっては利用している IP アドレスが変更される場合がある．

1.2 研究目的

既存の IP Geolocation サービスでは，IP アドレスの検索を行った場合，多くがサーバがある場所か登録が無い場合は「皇居」など日本の代表的な物をさしてしまう，さらに，公衆無線 LAN などから利用すると，短期間に IP アドレスが変動してしまう問題があることがある．そこで，本研究ではユーザが IP アドレス情報を登録できるシステム Prefecture maP Internet Protocol(以下，PPIP[*1]とする) を開発し，動的に変更される IP アドレスの調査を目的とする．

1.3 IP Geolocation

ユーザの位置情報を扱う技術として Geolocation がある．Google 検索ではこの Geolocation 技術を活用して「飲食店」などと検索するとユーザが検索した現在地の付近にある飲食店が検索結果の上位に表示される．これは Google が検索結果にユーザの位置情報を利用しているから，IP Geolocation とは IP アドレスからサーバの地理的位置情報を提供する技術のことを表す．

1.4 論文構成

本論文の構成は次の通りである．
まず，第 2 章で Prefecture maP Internet Protocol のシステムについて，3 章で研究室のサーバのアクセスログ調査を行う．PPIP の問題点から 4 章で動的 IP アドレスの調査を行い，結果や考察から今後の課題をの

[*1] PPIP　http://windy.mind.meiji.ac.jp/ atumi/ppip/site/HdbClick.php

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 システム開発担当を表 2.1 に示す．

2.2　使用 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 章 2.2 を参照．

2.3　TLS 公開鍵証明書

IP アドレスによる位置情報システムの開発と評価 (2)，3 章を参照．

表 2.1　PPIP システム開発担当

名前	担当
高橋	Google Maps JavaScript API[6],Geolocation API[8],TLS 公開鍵証明書の導入
笹	Google Maps JavaScript API[6],Google Places API[7]
厚見	データベースの設計・作成 [5]，データのやりとり [5]

図 2.1　登録システム画面

2.4　登録システム

ユーザは PPIP の登録画面にアクセスすると自身が使用している IP アドレスと位置情報を得ることができる．この 2 つに加えて都道府県名・位置情報の詳細・匿名の情報提供者名を入力して登録する．また，電波状況が悪い場合や位置情報の取得に時間がかかる場合，ユーザが位置情報取得を許可しなかった場合でも検索フォームにより自分のいる位置を検索して登録することや，GoogleMap を直接タップもしくはクリックして登録することができる．登録システム画面を図 2.1 に表す．

登録システムでは IP アドレスと現在地の緯度経度に加え，都道府県名，現在地情報，匿名の名前を入力し

図 2.2　現在地検索実行例

て登録する．また，位置情報の取得に失敗した場合や位置情報の取得を不許可にした場合に検索フォームを用意し，現在地周辺の地名を検索し，GoogleMap 上の現在地を指定してもらうことで緯度経度の情報を取得できるようにしている．図 2.2 に現在地検索システムの実行例を示す．現在地検索を行うと GooglMap が検索結果の地図を示し，ポイを立てる．

現在地検索では検索したい地名を検索フォームに入力し，現在地検索ボタンを押すと図 2.2 の GoogleMap 上で検索した地名に移動し，該当する場所にピンが立つ．このピンをクリック，もしくはタブレット機ならタップすることにより緯度，経度，都道府県，現在地名を自動的に入力し，登録をスムーズに行うことができる．

登録内容例を表 2.2 に，データベース設計を表 2.3 に示す．実際にデータベースに格納されるのは登録内容

表 2.2　登録データ例

id	IP アドレス	緯度	経度	都道府県名	所在地名
1	133.26.34.22	35.706962	139.659547	東京都	明治大学
2	116.82.xxx.xxx	36.060xxxx	139.509xxxx	埼玉県	xx 駅
3	180.43.xxx.xxx	35.4222xxxx	139.463xxxx	神奈川県	店名 xx 店舗

表 2.3　データベース構造

	id	IP アドレス	緯度	経度	都道府県名	所在地名	時刻	名前
データ型	int(11)	varchar(15)	double	double	varchar(6)	varchar(20)	timestamp	varchar(20)
変数名	id	ipaddress	lat	lng	prefec	name	timestamp	nick

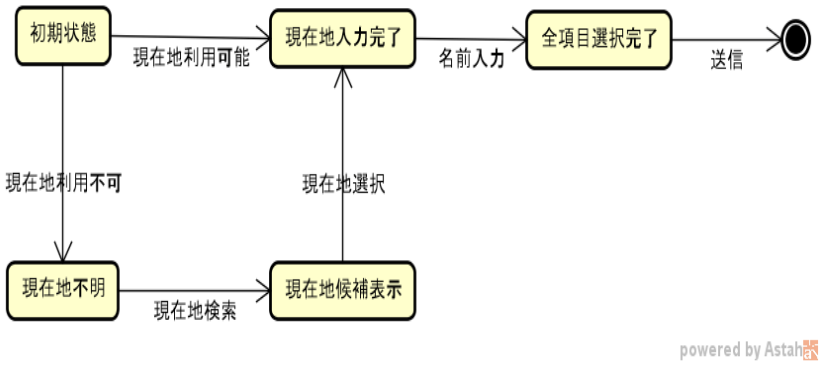

図 2.3　登録システムフロー図

に加えて id，時刻である．これら二つは登録時に自動的にデータベースに格納される．

登録システムのフロー図を図 2.3 に表す．

PPIP 登録システムで行っている各処理について記述する．トップ画面の地図を表示するシステムでは Google Map API を利用して地図を表示している．次に現在地利用を行わなかった場合，Google place API を使ってキーワードを検索すると検索した場所を表示し，現在地をクリックすると，緯度，経度，都道府県，所在地名を一部例外を除いて自動で入力されるようにしている．例外は Google 側で入っているワードの設定が異なる場合に発生する．手動でニックネームを入力し，送信ボタンを押す．

次に現在地検索の利用を行った場合，緯度，経度が自動で入力され，Google Map が現在地を表示しているので，都道府県名，所在地名（現在地に最も近い目印，例　駅名など），ニックネームを入力し，送信ボタンを押す．

送信ボタンを押すとデータベースに登録される．

図 2.4　検索システム図

2.5　検索システム

IP アドレスからその位置情報を提示する．IP アドレスが登録済みの場合は位置情報を提示する．未登録の場合は，MaxMind が提供するデータベースの情報を参照して提示する．検索システム画面を図 2.4 に示す．

PPIP では，検索フォームに IP アドレスを入力し，検索ボタンを押すと PPIP データベースにアクセスし，IP アドレスを検索し，情報を取得，IP アドレスの情報を取得すると，その IP アドレスの情報を画面に表示し，GoogleMap にその情報をマッピングする．その図を 2.5 に示す．

検索システムのフロー図を図 2.6 に示す．

IPアドレスを検索

結果

ipアドレス:133.26.35.40
緯度:35.706962
経度:139.659547
県名:東京都
店舗名:明治大学中野キャンパス

戻る

図 2.5　検索結果

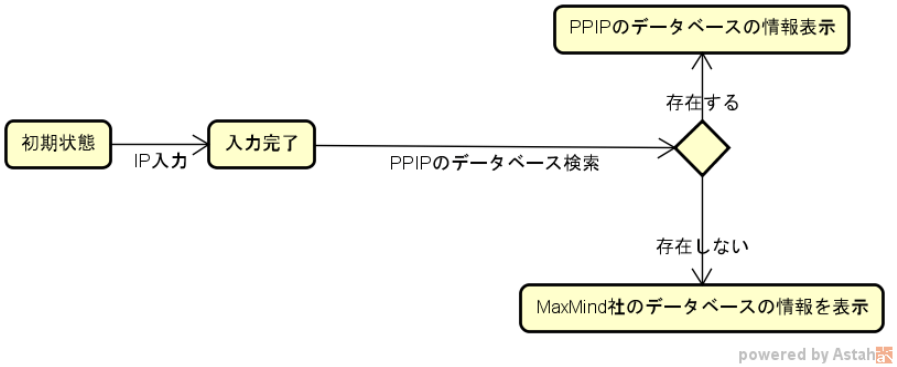

図 2.6　検索システムフロー図

表 2.4　PPIP データベース

	IP アドレス	緯度	経度	都道府県名	現在地名	登録日時
	”210.160.37.174”	”35.68”	”139.70”	”東京都”	”ルミネ新宿”	”2016-11-15 21:24:32”
	”153.209.194.229”	”35.68”	”139.70”	”東京都”	”東急ハンズ 新宿店”	”2016-11-15 21:24:32”
	”101.102.202.77”	”35.69”	”139.70”	”東京都”	”マクドナルド　西武新宿駅前店”	”2016-11-15 21:24:32”
	”103.5.140.140”	”35.68”	”139.69”	”東京都”	”新宿ミロード”	”2016-11-15 21:24:32”
	”133.26.35.86”	”35.70”	”139.65”	”東京都”	”明治大学中野キャンパス”	”2016-11-15 21:24:32”
	”183.77.251.13”	”35.67”	”139.70”	”東京都”	”クエストコート原宿”	”2016-11-16 10:51:45”
	”49.106.193.77”	”35.66”	”139.38”	”東京都”	”みずほ銀行日野支店”	”2016-11-17 13:28:45”
	”118.103.63.146”	”35.70”	”139.66”	”東京都”	”マクドナルド 中野セントラルパーク店”	”2016-11-17 16:23:44”
	”153.159.161.106”	”35.69”	”139.77”	”東京都”	”秋葉原駅”	”2016-11-24 14:16:36”
	”103.5.140.167”	”35.73”	”139.70”	”東京都”	”スターバックス コーヒー池袋西口店”	”2016-11-24 13:44:10”
	”210.160.37.43”	”35.72”	”139.70”	”東京都”	”東京芸術劇場”	”2016-11-24 13:44:20”

2.6　データ収集実験

2.6.1　実験方法

PPIP データセットの作成のため IP アドレスと位置情報を収集した．実験は同大学の学生とその友人らを対象に，期間は 2016 年 11 月 15 日から 2016 年 12 月 16 日の間に行った．被験者は自宅や外出先の施設内で利用できる Wi-Fi を利用してもらい，PPIP の登録サイトにアクセスしてもらい，IP アドレスと位置情報の登録をする．

2.6.2　実験結果

収集したデータ例を表 3.2 に示す．表 3.2 にあるものの他に登録順に割り当てられる ID とニックネームがあるが省略している．データの形式は IP アドレス，都道府県名，現在地名，ニックネームが文字列．緯度，経度が double 型．登録日時が timestamp 型である．
45 名の方に参加してもらい 165 件のデータが集まった．

第 3 章

研究室アクセスログの調査

3.1　調査概要

研究室に置いてあるサーバ上のどのサイトが一番多く参照されているのか、またどこからどれだけのアクセスがあるのかを調査する．サーバの情報を表 3.1 に示す．

3.2　内容

下記の課題それぞれについて調査を行った．

1.　研究室で一番参照数の多いサイトの抽出．
2.　日ごとに時間帯毎のアクセス数，通信料についてまとめる．
3.　/8 ブロック毎に IP アドレスの数を求める．

使用したアクセスログの例をログ 3.1 に，それぞれが何を表しているのかを表に示す．使用するログの中身は ip アドレス，アクセスした時間、リクエスト記録，ステータスコード，通信のバイト数，リンク元 url，ブラウザの種類や端末情報 を用いた．

ログ 3.1　アクセスログの例

```
xxx.xxx.xxx - - [31/Jul/2016:19:23:52 +0900]
"GET http://windy.mind.meiji.ac.jp/kiknlab2014/index.html"
"Mozilla/5.0 (iPhone; CPU iPhone OS 9_2 like Mac OS X)
AppleWebKit/601.1.46 (KHTML, like Gecko)
Mobile/13C75 Twitter for iPhone"
```

表 3.1　サーバ情報

IP アドレス	ドメイン	コンテンツ	OS
133.26.34.230	windy.mind.meiji.ac.jp	菊池研究室ホームページ	CentOS

表 3.2　参照数

順位	回数	サイト名
1	79172	電気第 5 研究室合宿用掲示板
2	47477	匿名加工班の評価プラットフォーム
3	3744	”-”:
4	1152	第 16 回情報ネット

表 3.3　時間帯毎の参照数 (回数)

日月	8/25	8/26	8/26	8/24	8/27	8/24	8/24	8/24	8/27
時刻	22	22	23	20	00	19	18	17	15
回数	1820	1800	1789	1780	1776	1767	1757	1744	1738

表 3.4　/8 ブロック毎の回数

回数	先頭 1 ブロック	所属 RIR
79097	91	RIPENCC
21152	153	APNIC
15671	119	APNIC
10794	133	APNIC
2350	113	APNIC
1025	180	APNIC

3.3　実験結果

3.3.1　1. 研究室で参照数の多い上位 4 サイトの抽出.

一番多く参照されたサイトと回数を表 3.2 に示す.

3.3.2　2. 日ごとに時間帯毎のアクセス数，通信量

日ごとのアクセス数について図**??**に，時間帯毎の参照数について表 3.3 に示す．調査した期間が短いため，ここからでは特徴を見つけることができなかった.

3.3.3　3./8 ブロック毎に IP アドレスの数

/8 ブロックとは，全 IP アドレス約 46 億個を 256 個に分割したもの．(XXX.XXX.XXX.XXX の先頭 1 ブロック毎) この 256 個のアドレスをそれぞれの地域インターネットレジスト リ (RIR) に配布している．一つのブロックには約 1670 万のアドレスが内包されている．RIR は全部で 5 つあり，それぞれ AfriNIC(アフリカ地域)，APNIC(アジア 太平洋地域)，ARIN(北米地域)，LACNIC(中南米地域)，RIPE NCC(欧州 地域) となっている．調査で使用したアクセスログを/8 ブロック毎に分類した結果の内上位 6 個を表 3.4 に示す.

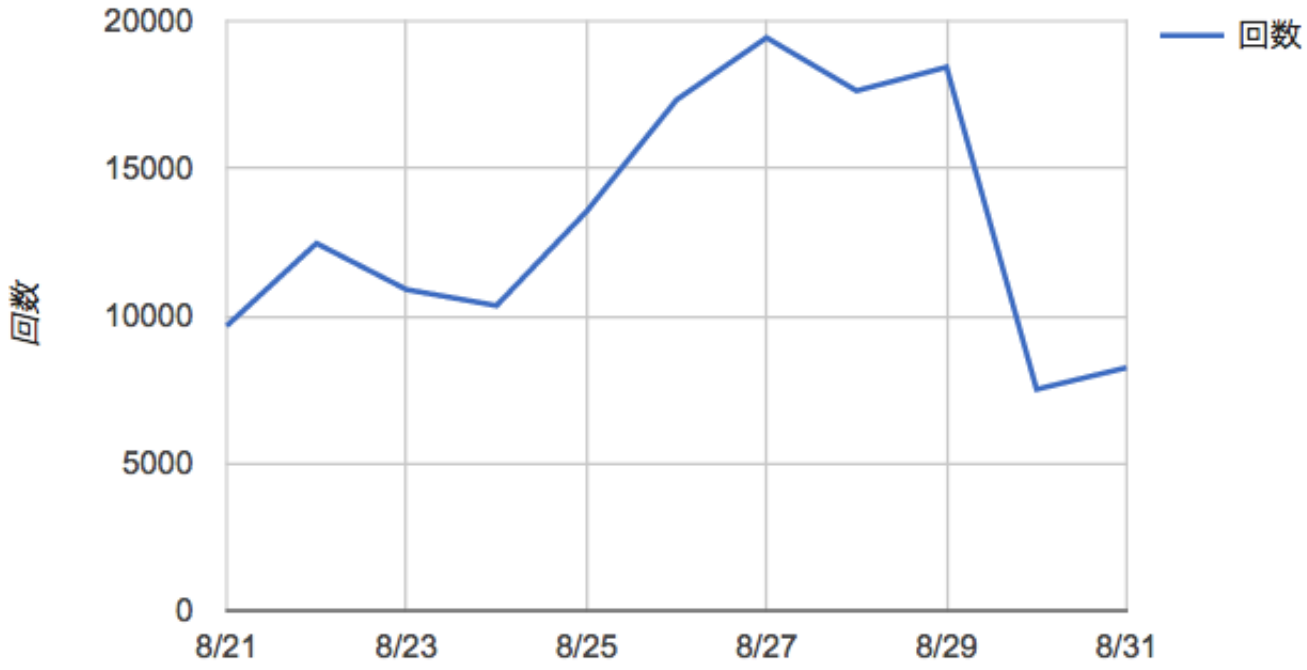

図 3.1　日ごとのアクセス数

表 3.5　/8 ブロック毎の回数

個数	所属 RIR
51	APNIC
45	ARIN
27	RIPE NCC
6	LACNIC
3	AFRINIC
1	UKGD
1	PSI
1	DuPont

ブロック毎の回数だとアジア圏 (APNIC) ではなく欧州地域 (RIPENCC) が多かったため，所属 RIR 毎に分類してみる．その結果を表 3.5 に示す．この結果から一番多いのはアジア圏，二番目に多いのは北米地域であることがわかり，欧州地域は三番目にあり，ブロック毎に分けた時のアクセス数から特定の IP アドレスからの回数が多いのであろうことがわかる．さらに下位三つはどの RIR にも属さず，企業毎に管理していることがわかった．

3.4　まとめ

アクセスログの調査を行った．一番参照数が多かったのは掲示板でアクセスが多いというよりもサーバ上のサイトにアクセスされるたびに更新される使用だったため一番多く回数があった．日ごとに時間帯毎のアクセス数，通信料についてまとめてみたが調査した日数が短く，これといった特徴は見えなかった．/8 ブロック毎に IP アドレスの数を求めてみた結果，やはりアジア圏からのアクセス数が一番多く，外国の企業が大規模なネットワークを所有していることもわかった．

第 4 章

動的 IP アドレスの調査

4.1　調査概要

3 章で行ったデータ収集実験のでデータを見ていると，同じ場所なのに別の IP アドレスが登録されていることがわかった．従って IP アドレスの変更が確認された Wi-Fi にある程度の間隔でアクセスし，データの収集を行った．

4.2　DHCP

Dynamic Host Configuration Protocol(以下,DHCP とする) とは，インターネットに接続するのに必要な IP アドレスを指定された範囲内で重複無く自動的に割り当てる機能をいう．インターネットを利用する時，インターネットサービスプロバイダの DHCP サーバがその度自動的に IP アドレスを割り当てる．そのため，インターネットに接続する度に異なる IP アドレスが割り当てられることがある．この IP アドレスを「動的 IP アドレス」という．

表 4.1　IP アドレス割当履歴

	IP アドレス	日付	時刻	認証
1	103.5.140.161	2016-12-17	15:44:48	有
2	103.5.140.141	2016-12-17	17:13:31	有
3	103.5.140.141	2016-12-17	17:14:04	無
4	103.5.140.175	2016-12-17	22:08:03	有
5	103.5.140.147	2016-12-18	10:54:43	有
6	103.5.140.166	2016-12-18	16:07:27	有
7	103.5.140.147	2016-12-18	16:08:54	無
8	103.5.140.137	2016-12-18	17:01:59	無
9	103.5.140.137	2016-12-18	17:02:54	無
10	103.5.140.149	2016-12-18	19:39:39	無
11	103.5.140.188	2016-12-18	21:21:07	無

表 4.2　IP アドレス割当履歴

	IP アドレス	日付	時刻	認証
1	118.103.63.145	2016-12-17	15:43:22	有
2	118.103.63.147	2016-12-18	10:53:49	有
3	118.103.63.149	2016-12-18	16:09:12	有
4	118.103.63.145	2016-12-18	17:03:10	無
5	118.103.63.158	2016-12-18	19:40:27	有

4.3　データ収集実験

IP アドレスの変化が確認された同一の Wi-Fi に，時間毎にアクセスして収集したデータから IP アドレスの変更間隔を調査する．神奈川県にあるスターバックスとマクドナルドの 2 店舗の FREEWi-Fi から１２月１７日と１２月１８日の二日間かけて時間毎にデータを収集した．

4.4　結果

IP アドレス割当履歴を表 4.1 と表 4.2 に示す．IP アドレス，日付，時刻，認証の有無のデータを使用した．Wi-Fi に接続する際に認証手続きがある時に認証している．実験をした FREE Wi-Fi の認証は二店舗ともユーザがするクリック認証である．表 4.1 と表 4.2 から接続する度に IP アドレスが変わっている．しかし，表 4.1 の 8 と表 4.1 の 9 を見ると極短い時間の間隔だと同じ IP アドレスを割り当てている．

表 4.3　PPIP データベース

	IP アドレス	日付	時刻	都道府県名	現在地名
1	103.5.140.137	2016-11-15	21:24:32	埼玉県	スターバックス
2	103.5.140.175	2016-11-16	11:23:40	神奈川県	スターバックス
3	118.103.63.158	2016-11-24	13:43:24	東京都	マクドナルド

表 4.4　PPIP データベース

	IP アドレス	日付	時刻	都道府県名	現在地名
1	103.5.140.137	2016-12-18	17:01:59	神奈川県	スターバックス
2	103.5.140.175	2016-12-17	22:08:03	神奈川県	スターバックス
3	118.103.63.158	2016-12-18	19:40:27	神奈川県	マクドナルド

4.5　考察

表 4.2 を見ると，認証が無いにも関わらず表 4.2 の 3 と表 4.2 の 4 では IP アドレスが変更されている．実験を行った FREE Wi-Fi は連続利用可能時間が 1 時間と定められていることから，表 4.2 の 3 と表 4.2 の 4 の時刻を見ると利用時間を越えて接続が切れた訳ではないと考えられる，そのため FREE Wi-Fi の接続を切っていた間に別の端末に IP アドレスが割り当てられてしまったと考えられる．しかし，表 4.1 の 2 と表 4.1 の 3，表 4.1 の 6 と表 4.1 の 7 を見てみると，両方とも前回接続時から 2 分以内に接続し，認証が無いにも関わらず IP アドレスに変更があるものとないものが確認された，さらに表 4.1 と表 4.2 の IP アドレスを見てみるとどちらも店舗毎に利用できる IP アドレスが一定の範囲以内にある特徴がある．表 4.3 に PPIP に登録されているデータの内，同じ IP アドレスが別の場所に登録されているものを示す．

表 4.4 は動的 IP アドレスの調査のため二日間収集したデータの一部を表す．収集したデータを PPIP のデータを比較すると所在地名が別の場所なのに同じ IP アドレスが割り当てられていることがわかる，このことからプロバイダーが一定の IP プールを所持しており，そこから各店舗に IP アドレスを割り当てていっているのではないかと考えられる，実験 2 で収集した IP アドレスを whois[3] で調べると，プロバイダーは 103.5.140.0 - 103.5.140.255 の範囲の IP アドレスを所持し，実験で収集した IP アドレスは全てこの範囲に含まれていた．

第 5 章

おわりに

既存の IP Geolocation システムは精度が低いため，IP アドレスや位置情報を登録するシステムを開発した，動的に変更する IP アドレスの調査を行った．本実験では FREE Wi-Fi に接続してから 1 時間未満で IP アドレスの変更や，2 分未満での接続での IP アドレスの変更が確認されたが，変更間隔正確な時間や条件を特定するには至らなかった，さらに同一の IP アドレスが別の場所で観測できたことが新たにわかった．今後の課題としては，検索システムにおいて動的に変更する IP アドレスの扱い，IP アドレスの重複，また IP が変更された場合に自動で取得する方法の検討がある．

謝辞

本研究にあたり，ご指導頂いた卒業論文指導教員の菊池浩明教授に感謝致します．また PPIP データセットを作るにあたり，実験に協力していただいた皆様に感謝致します．

参考文献

[1] Geo IP Tool（https://geoiptool.com/ 2016 年 10 月年参照）

[2] GeoIP MaxMind (https://www.maxmind.com/ja/geoipdemo 2016 年 10 月参照)

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

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

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

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

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

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

[9] Syncer (https://syncer.jp/how-to-use-geolocation-api 2016 年 10 月参照)

[10] 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

[11] パケットキャプチャ入門-第 3 版-LAN アナライザ Wireshark 活用術-竹下-恵.（2016 年 7 月参照）

付録 A

DbD 攻撃の検知

A.1　研究背景

普段何気なく使っていたサイトが気づかないうちに悪性サイトに代わり，ウィルスに感染してしまう事が増えている．Drive-by-Download(以下，DbD) 攻撃とは web を利用してマルウェアを拡散する攻撃であり，ユーザは攻撃を仕掛けられた web ページにアクセスしただけで感染する．日々多様化、高度化する DbD 攻撃，難読化 Javascript の発達などにより検出し，解析し，対策をとるまでの流れに時間かかってしまう問題がある．

A.2　研究目的

DbD 攻撃通信の特徴を抽出し，それを用いて攻撃通信を検出できるようにする．

A.3　決定木

決定木とは，分析手法の一つで要因を分析し，その分析結果が樹木状のモデルで表されるものである．決定木の例を図 A.1 に表す．四角で囲まれている部品それぞれをノードという．一番上のノードをルートといい，上から下に向けて一方向に分析を行う．決定木の分岐を先をブランチという．たとえば「男性」から下の 4 つをノード「男性」のブランチといい，ブランチの末端部分をターミナルノードという．図 A.1 には 5 個のターミナルノードがあると言える．文献 [3] を参照.

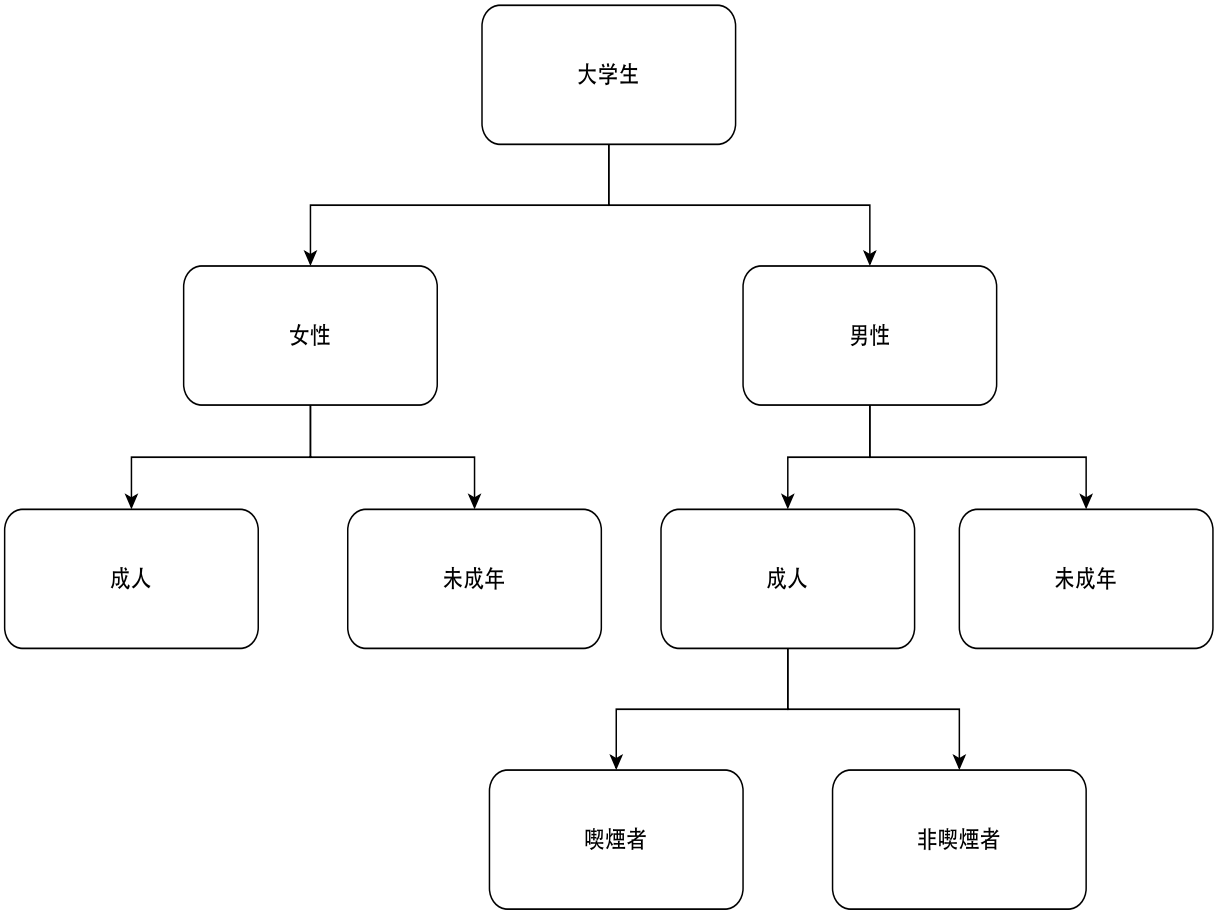

図 A.1　決定木

A.4　研究概要

D3M データセットの悪性通信から特徴量を抽出、正規の通信から特徴量を抽出．抽出する特徴量は「Drive-by-Download 攻撃通信可視化システム [1]」であげられている物を用いる．使用する特徴量は Javascript ファイル，JAR ファイル，PDF ファイル，EXE ファイルそれぞれの byte 数，リダイレクトに用いられる iframe，301，302 の回数，難読化処理をするための unescape 関数の有無などを使う．まず D3M 2012 にある悪性通信 51 セッション分と正常通信として AlexaTop サイトの上位 31 件から計 82 セッション分を用意する．それらから特徴量を抽出し，R の決定木を使い，検知率の評価を行う．その後に悪性通信のセッション数を正常通信のセッション数を増やして検出を行っていく．

表 A.1　特徴量_1

	packetSUM	Jsbyte	Jarbyte	Exebyte	Pdfbyte	Flashbyte
amazon	191510	10037	0	0	0	1208
2012_0328_1	3553	0	1314	0	1334	0

表 A.2　特徴量_2

	times301	times302	jsDLend	PDFDLstart	hostsum	unescape	D3M
amazon	14	0	7.992682	1	21	no	no
2012_0328_1	0	0	-1	20.109125	4	no	yes

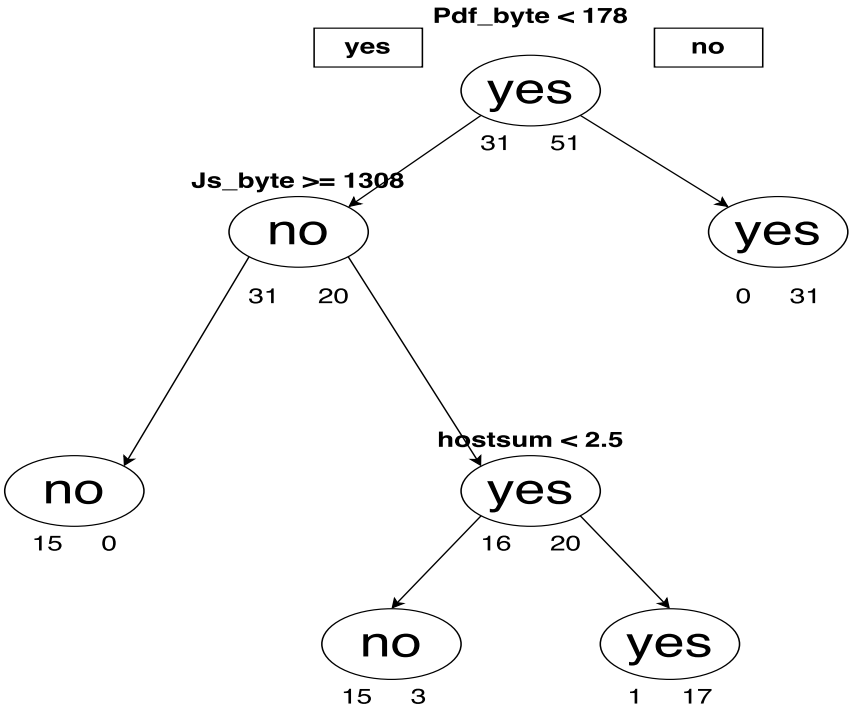

図 A.2　決定木

A.5　実験

用意した特徴量から決定木の作成を行う．使用した特徴量の例を表 A.1，A.2 に表す．D3M 変数が yes の物は D3M データセットに入っており，no の物は安全な通信のデータセットに入ってるものを示す．

A.6　実験結果

実験結果を図 A.2 に示す．誤検知した通信の例を表 A.3，A.4 に示す．

表 A.3　誤検知_1

	packetSUM	Jsbyte	Jarbyte	Exebyte	Pdfbyte	Flashbyte
taobao	1057	0	0	0	0	0
2012_0328_5	0	0	0	0	0	0
2012_0328_32	0	0	0	0	0	0
2012_0328_33	0	0	0	0	0	0

表 A.4　誤検知_2

	times301	times302	jsDLend	PDFDLstart	hostsum	unescape	D3M
taobao	0	2	1	1	4	no	no
2012_0328_5	1	0	1	1	2	no	yes
2012_0328_32	1	0	-1	1	4	no	yes
2012_0328_33	1	0	-1	1	2	no	yes

A.7　考察

全体としては 82 個の通信のうち 4 個が誤検知された．PDF，JAR ファイル，exe ファイルは DbD 攻撃通信のみに見られたが，今回の決定木の判別では PDF しか使われなかった．今後の課題としては，誤検知した通信の分析，特徴量の抜け漏れがないかの精査，SVM などの他の機械学習との比較があげられる．

参考文献

[1] 松本浩明, 石井啓介, 薄羽大樹, 菊池浩明,”Drive-by-Download 攻撃通信可視化システム”,CSS 2014,pp.9-16,2014.

[2] パケットキャプチャ実践技術ー Wireshark によるパケット解析　応用編.（2015 年参照）

[3] 豊 田 秀 樹 ,” データマイニング入門”　(http://www.tokyo-tosho.co.jp/books/ISBN978-4-489-02045-2.html,2015 年参照).