卒業論文　　2013年度（平成25年度）

マイクロブログを活用した列車運行状況モニタリング

慶應義塾大学 環境情報学部

氏名：深谷 哲史

担当教員

慶應義塾大学 環境情報学部

村井 純

徳田 英幸

楠本 博之

中村 修

高汐 一紀

重近 範行

Rodney D. Van Meter III

植原 啓介

三次 仁

中澤 仁

武田 圭史

平成25 年10 月17 日

卒業論文要旨 - 2009 年度 (平成 21 年度)

デジタル情報収集による ユーザ追跡とリスク分析と対策の提案

情報技術の発展に伴い，ネットワーク上に発信されるデジタル情報は容易に記録できるようになった．これによって，今まで単独ではユーザの個人情報とならなかった情報を複数組み合わせることで，ユーザプライバシが侵害される可能性がある．この問題に取り組むためには，ユーザが定常的に発信している情報を組み合わせた際に，どの程度までユーザプライバシが脅かされるのかを明確にして議論する必要がある．そして，ユーザプライバシを守るために，これまで個人情報と見られていなかったものも含めて，情報の収集と取り扱いに関するガイドラインを明確に取り決めなければならない．

本論文では，ユーザが無意識に発信している情報の収集によって，ユーザプライバシが侵害される可能性を提示する．個人情報になりうるユーザ情報は，情報収集者と対象になるユーザとのネットワークの上での関係によって取得できる範囲が変わり，リスクも変化する．そこで，一般的に取得可能であると見込まれる情報を 3 種類挙げ，それぞれの情報によって，ユーザのプロファイルを作成する手法を提示した．本論文でプロファイル作成に利用した情報は，パケットのヘッダ情報，ホスト資源共有に関する情報，Bluetooth デバイスの探索情報である．これら 3 つの情報は多くのユーザが定常的に発信しているため，収集が容易である．これらの情報によっては，ユーザを特定することができれば，ユーザのネットワークにおける行動履歴や，実際の生活時間や場所など，ユーザプライバシが脅かされる危険性がある．そして，提示した手法を実証するために，各情報を収集・解析するシステムを実装し，検証した結果，前述した 3 つの情報を利用してユーザのプロファイルが作成できることを確認した．

これらの成果に基づき，3 つの情報を利用してデータを収集するケースを想定し，ユーザのプライバシに対する影響を考察した．そして，ユーザのプライバシを保護するために，検証結果に基づいたガイドラインを提案した．

キーワード:
1. ネットワーク追跡, 2. フォレンジック, 3. セキュリティ, 4. ネットワーク監視,

慶應義塾大学 総合政策学部

上原 雄貴

Abstract of Bachelor's Thesis - Academic Year 2009

Risk Analysis and Countermeasures on User Tracking by Digital Information Surveillance

As computer networks have covered various places and population globally, users transmit various data in numerous occasions, both intentionally and unintentionally. As services that utilize the network increased, the chance of data transmitted on the network being accumulated and recorded has reached the significant level. Those individual data may not be considered as privacy information. However, as those control data has increased, it became possible to combine them and produce a single profile of a certain user. When the profiling become possible, the information that weren't considered as a privacy information then becomes a privacy information.

To ensure that the users' privacy aren't intruded, it is necessary to determine which information could lead the profiling of the user, and construct a guideline based on the study. This thesis clarifies the types of information that could be accumulated to profile a user, and how those information could be captured on the computer network.The method proposed in the thesis classifies collectors into three categories, and different methods of profiling is stated based on the characteristics of those categories. The information used for capturing a user's profile includes: packet header information, information used for sharing hosts' computing resources, and device discovery information for Bluetooth devices. The threats that could outcome from the profiling include: revealing users' activity history, discovering when the users are actively using the network, and determining actual location of the physical computer that is being a source of the information.The system for capturing and analyzing those information was developed to present that they could be a threat against privacy information. The result showed that both specifying an individual user and profiling the user's activities is possible based on the method presented in the thesis.

Based on the evaluation,we discussed cases of collecting these imformation and impact of users privacy. Additionally, the guidelines for handling those information is proposed, to ensure that the users' privacy are protected and secured.

Keywords :
1. Network Tracking, 2.Digital Forensics, 3.Internet Security, 4.Network Monitoring

Keio University, Faculty of Policy Management

Yuki Uehara

目 次

図 目 次

表 目 次

第1章　序論

1.1　電車遅延の現状と対策

1.2　ビッグデータの活用

1.3　本研究の目的

1.4　本論文の構成

第2章　背景

本章では，電車の遅延に関する人々への影響や鉄道会社が行っている公式の情報発信について述べ，そこに生じる問題点を示し，その解決のために本論文で提案するビッグデータ解析に関連した技術や研究について述べる．

2.1　電車

電車は多くの人に通勤，通学の手段として活用されている．しかし，遅延や運行見合わせなど問題点もある．本説では，電車の問題点と鉄道会社が行っている公式の対応について示す．

2.1.1　電車の問題点

日本において，多くの人が通勤，通学の手段として電車を活用している．電車の利点は短時間で長距離移動することができるという点である．また車と違い道路の渋滞もなく，時刻表通りに運行が行われるため利用者は正確な移動時間を考慮した上で利用することができる．そのため，様々な場面で様々な人に活用されている．国土交通省が公表している旅客の輸送機関別輸送量の図 2.1[1] によると，年間約 230 億人の日本人が交通の手段として電車を活用している．電車の利用人数は他の輸送機関よりも多く，日本において電車はとても重要な役割を果たしている．

しかし，問題点もある．時間に正確であるということから多くの人に活用されている電車だが，事故や整備点検などによって遅延や運行見合わせなどが生じてしまうことが多くある．東洋経済オンラインの記事 [2] の中で岩倉成志教授によると，電車の遅延による都区内への通勤にかかる社会的費用は年間 2180 億円にもなると推測されている．正確な数字とは言えないが，電車の利用者にとって電車の遅延がとても多くの損失を与えていると言える．

2.1.2　鉄道会社の遅延に対する対応

電車は遅延や運行見合わせなどによって，時刻表通りの運行を行えていないことが多々ある．遅延や運行見合わせに対して，各鉄道会社は様々な対応を行っている．リアルな対

分類 / 年度	輸送人員（単位：千人） 自動車	指数	鉄道	指数	うちJR（国鉄）	指数	旅客船	指数	航空	指数
昭和25	1,515,000	(5.0)	8,391,932	(47.7)	3,095,194	(43.9)	97,348	(57.3)	—	—
30	4,261,000	(15.0)	9,780,980	(55.6)	3,849,219	(54.6)	73,920	(43.5)	361	(1.4)
35	7,900,743	(27.3)	12,290,380	(69.9)	5,123,901	(72.7)	98,887	(58.2)	1,260	(4.9)
40	14,863,470	(52.3)	15,798,168	(89.8)	6,721,827	(95.4)	126,007	(74.2)	5,194	(20.4)
45	24,032,433	(84.6)	16,384,034	(93.2)	6,534,477	(92.7)	173,744	(102.3)	15,460	(60.7)
50	28,411,450	(100.0)	17,587,925	(100.0)	7,048,013	(100.0)	169,864	(100.0)	25,467	(100.0)
55	33,515,233	(118.0)	18,044,962	(102.4)	6,824,817	(96.8)	159,751	(94.0)	40,427	(158.7)
60	34,678,904	(122.1)	18,989,649	(108.0)	6,943,358	(98.5)	153,477	(90.4)	43,777	(171.9)
平成 2	55,767,427	(196.3)	22,029,909	(125.3)	8,357,583	(118.6)	162,600	(95.7)	65,252	(256.2)
6	59,934,869	(211.0)	22,679,748	(128.9)	8,883,691	(126.0)	150,866	(88.8)	74,547	(292.7)
7	61,271,653	(215.7)	22,708,819	(129.1)	8,982,280	(127.4)	148,828	(87.6)	78,101	(306.7)
8	61,542,541	(216.6)	22,673,706	(128.9)	8,997,038	(127.6)	148,107	(87.2)	82,131	(322.5)
9	62,199,844	(218.9)	22,325,628	(126.9)	8,859,635	(125.7)	144,897	(85.3)	85,555	(335.9)
10	61,838,994	(217.7)	22,068,065	(125.5)	8,748,331	(124.1)	127,665	(75.2)	87,910	(345.2)
11	62,046,830	(218.3)	21,809,976	(124.0)	8,701,483	(123.5)	120,091	(70.7)	91,588	(359.6)
12	62,841,306	(221.2)	21,705,687	(123.4)	8,654,436	(122.8)	110,128	(64.8)	92,873	(364.7)
13	64,590,143	(227.3)	21,779,603	(123.8)	8,634,327	(122.5)	111,550	(65.7)	94,579	(371.4)
14	65,480,675	(230.5)	21,647,202	(123.1)	8,586,192	(121.8)	108,846	(64.1)	96,662	(379.6)
15	65,933,252	(232.1)	21,840,622	(124.2)	8,652,606	(122.8)	107,288	(63.1)	95,487	(374.9)
16	65,990,529	(232.3)	21,810,623	(124.0)	8,616,982	(122.3)	100,872	(59.4)	93,739	(368.1)
17	65,946,689	(232.1)	22,614,234	(128.6)	8,683,855	(123.2)	103,175	(60.8)	94,490	(371.0)
18	65,943,252	(232.1)	22,688,880	(129.0)	8,778,188	(124.5)	99,200	(58.3)	96,971	(380.8)
19	66,908,896	(234.5)	22,921,594	(130.3)	8,987,947	(127.5)	100,800	(59.3)	94,849	(372.4)
20	66,774,143	(235.0)	23,071,018	(131.2)	8,984,940	(127.5)	99,000	(58.3)	90,662	(356.0)
21	66,599,647	(234.4)	22,984,742	(130.7)	8,840,512	(125.4)	92,200	(54.3)	83,872	(329.3)
22	6,241,395	(22.0)	23,080,111	(131.2)	8,819,053	(125.1)	85,000	(50.0)	82,194	(322.7)

※自動車は、平成22年の東日本大震災の影響のため、22年度の数値には北海道運輸局及び東北運輸局管内の3月の数値は含まない。

図 2.1: 日本の輸送機関別輸送人員数

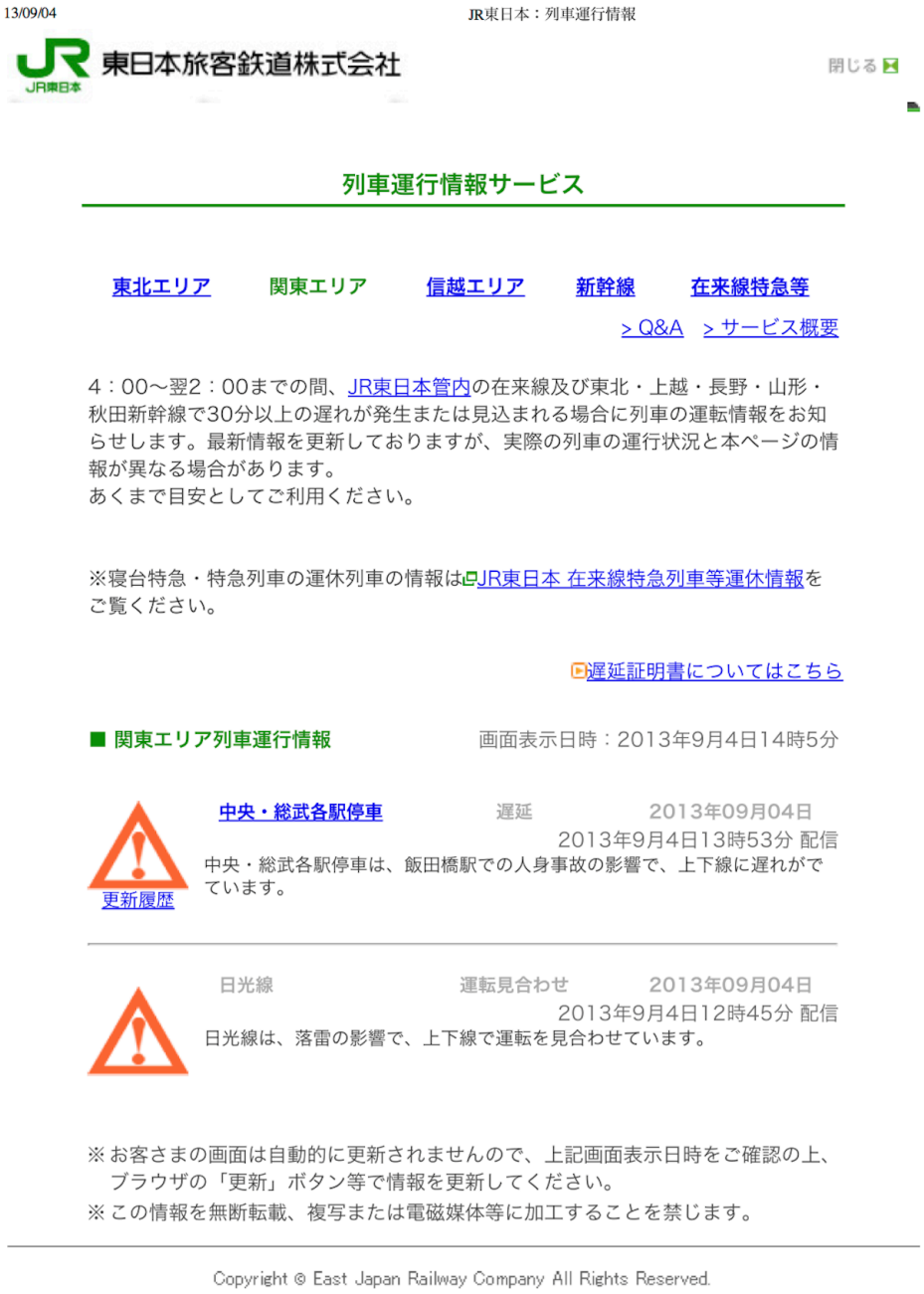
応としては，電工掲示板に状況の表示や駅で遅延証明書の配布などが行われている．リアルな対応はその場に行かなければ電車の運行状況を知ることができないため，利用者にとってはとても都合が悪い．そのため，駅に行かずに電車の運行状況を確認できるように各鉄道会社はオンラインで情報を発信している．オンラインでの情報発信の仕方を以下に挙げる．

(1) 公式サイトによる列車運行情報提供ページ

各鉄道会社はそれぞれ Web サイトを持っていることが多い．それらの Web サイトでは企業情報，ニュース，特急チケットの予約など様々の情報が発信されている．それらの情報の 1 つとして，列車の運行状況を発信しているページが存在する．図 2.2[3] のように駅に行かずに列車の運行情報を知ることができる．このページをうまく活用すれば，運休している路線を回避しタクシーを使うなど駅に行かずに最善の経路を選択することができる．しかし、問題点もある．公式情報ということもあり遅延や運行見合わせなどが発生してから配信まで時間がかかるという点である．そのため，情報の信用性は高いがリアルタイムに乏しい．

(2) Twitter 公式アカウント

Twitter に公式アカウントを持っている鉄道会社もある．そのアカウントをフォローすることによって，Twitter をやっているユーザは Twitter を見ている中で電車の運行情報を得ることができる．2.3[4] しかし，問題点もある．公式アカウントなので情報も公式のものなので，こちらも投稿までに時間がかかります，また，投稿される情報は実際に発生している遅延よりも少ない．


13/09/04 JR東日本：列車運行情報

東日本旅客鉄道株式会社

閉じる

列車運行情報サービス

東北エリア 関東エリア 信越エリア 新幹線 在来線特急等

> Q&A > サービス概要

4：00～翌2：00までの間、JR東日本管内の在来線及び東北・上越・長野・山形・秋田新幹線で30分以上の遅れが発生または見込まれる場合に列車の運転情報をお知らせします。最新情報を更新しておりますが、実際の列車の運行状況と本ページの情報が異なる場合があります。
あくまで目安としてご利用ください。

※寝台特急・特急列車の運休列車の情報はJR東日本 在来線特急列車等運休情報をご覧ください。

遅延証明書についてはこちら

■ 関東エリア列車運行情報 画面表示日時：2013年9月4日14時5分

中央・総武各駅停車 遅延 2013年09月04日
2013年9月4日13時53分 配信
中央・総武各駅停車は、飯田橋駅での人身事故の影響で、上下線に遅れがでています。
更新履歴

日光線 運転見合わせ 2013年09月04日
2013年9月4日12時45分 配信
日光線は、落雷の影響で、上下線で運転を見合わせています。

※お客さまの画面は自動的に更新されませんので、上記画面表示日時をご確認の上、ブラウザの「更新」ボタン等で情報を更新してください。
※この情報を無断転載、複写または電磁媒体等に加工することを禁じます。

Copyright © East Japan Railway Company All Rights Reserved.

traininfo.jreast.co.jp/train_info/kanto.aspx 1/1

図 2.2: JR 列車運行情報サービスページ

図 2.3: 小田急線 Twitter アカウントページ

2.2 SNS

SNS とは，ソーシャルネットワーキングサービス（Social Networking Service）の略である．社会的ネットワークを Web 上に構築することできるがサービス及びサイトである．基本的な機能として，プロフィール機能，メッセージ機能，ユーザ間における相互リンク機能と検索機能，ブログ機能，コミュニティ機能などがある．モバイル端末の普及によって，手軽に利用することが可能になり多くの人に活用されている．世界では Facebook, Twitter，Google+，日本では Mobage，GREE，Ameba，mixi などがあり，多くの人に利用されている．

2.2.1 ソーシャルセンサとしてのSNS

世界中のSNSユーザは日々，SNSを通して様々の情報を発信している．自分や自分の周りの状況，写真，読んだニュースの記事などを投稿することによって，SNS上の他のユーザに向けて情報を拡散している．またSNSの情報発信・拡散スピードはテレビ，ラジオ，新聞，雑誌・書籍などに比べ圧倒的に早い．This just in…News no longer breaks, in Tweetsはある ITコンサルタントがウサマ・ビン・ラディンが殺害された際，その一部始終をTwitterにリアルタイムに投稿していたことを記事にしたものである．従来であれば，記者が本人に取材し記事にして既存のメディアを通してニュースとして全世界に公開されるはずであった．しかし，Twitterを用いてリアルタイムに情報発信・拡散されたことによって，バラク・オバマ米大統領緊急声明を発表しウサマ・ビン・ラディンを殺害したことを明らかにする前に多くの人がそのことを知っていた．この事実はTwitterなどのSNSがリアルタイム性の高い情報発信媒体であることが言える出来事である．東日本大震災ではTwitterやfacebookが知人や家族の安否，地震の被害状況，電車の運行状況についての情報収集の手段としてとても重要な役割を果たした．近年では，人を物理センサと同様の機能を持つ一種のセンサと考え，ソーシャルメディアを活用してリアルタイムに実世界を観測するという考えが生まれきている．

2.3 ビッグデータ

ビッグデータとは，従来のデータベース管理ツールやデータ処理アプリケーションでは処理するのが困難なほど大量なデータ集合のことである．

2.3.1 ビッグデータの特徴

ビッグデータの特徴として3Vという言葉がある．3Vとは，ビッグデータの容量（Value），種類（Variety），頻度（Velocity）の3つの特徴のことである．以下でそれぞれの特徴について説明する．

- 容量（Volume）
  近年，モバイル端末の普及に伴い多くの人がインターネットを使用するようになった．そのため，インターネットに蓄積されるデータは膨大になってきている．世界最大級のSNSであるfacebookは2012年8月時点で500TBのデータを蓄積している．Twitterは2011年10月時点で1日に2億5000万ツイートを突破している．140文字の個々のツイートのデータ量は約200バイトなので，Twitterは1日に約8テラバイトものデータを生み出しているということになる．また，Googleは2008年時点で1日に20ペタバイトものデータを処理している．数年前に比べて，扱うデータが膨大になってきているという点がビッグデータのVolumeという特徴である．

- 種類（Variety）
  数年前に比べて，インターネットに様々な種類のデータが蓄積されるようになって

きている．データの種類は Web のログデータ，テキストデータ，画像，動画，携帯電話やタブレット端末の GPS（Global Positioning System）など以前は破棄されていたようなデータも蓄積されるようになってきている．特に近年急増してきているのが，インターネット上のテキストデータ，位置情報，アクセスログ，センサデータ，動画など従来の主流であったリレーショナル・データベースでは扱うことが困難な非構造データである．以前からも非構造データは存在し，蓄積されていた．しかし現在はただ単に蓄積するだけではなく，分析することによって有用な知見を得ようという点がビッグデータの Variety という特徴である．

- 頻度（Velocity）
  インターネット上に蓄積されるデータの発生頻度も以前に比べ，圧倒的に増えている．全国のコンビニエンスストアで発生する POS（Point Of Sales）データ，EC サイトでユーザがアクセスするたびに発生する Web のクリックストリームデータ，Twitter に投稿されるテキストデータ，監視カメラの動画，全国の道路に設置されている道路の渋滞検知センサや放射線を測定するセンサなどのセンサデータ，Suica や PASMO などの交通系の IC カードから生み出される乗車履歴データや電子マネーの決済履歴データである．このように 365 日 24 時間大量のデータを生み出し続けているという点がビッグデータの Velocity という特徴である．

2.3.2 ビッグデータを支える技術

ビッグデータが話題になったのは単にデータの量，種類，頻度が増えただけではない．ビッグデータを汎用品のサーバを用いて蓄積し，高速に処理することができるオープンソースのソフトウェア技術が生み出されたことが大きな要因の 1 つである．以下でビッグデータが話題となった要因の技術の説明をする．

- Hadoop
  Hadoop とは，Apache が開発しオープンソースとして公開している大規模分散処理フレームワークである．Hadoop はアプリケーションが数千ノードおよびペタバイト級のデータを処理することが可能である．Hadoop は Google が開発した MapReduce, Google File System（GFS），Big Table をもとに開発された．Hadoop は MapReduce, Hadoop Distributed File System（HDFS），HBase から構成される．MapReduce とは，膨大なデータセットに対して並列処理可能な問題を Map ステップ，Reduce ステップで処理をする処理方式のことである．MapReduce で複数台のサーバに処理を分散させることによって，1 台では何日もかかっていた処理を数時間で処理することが可能になった．

- NoSQL
  NoSQL とは，Not only SQL の略である．従来からデータ管理として，リレーショナルデータベース管理システム（RDBMS）が活用されている．RDBMS は，SQL という標準言語によってデータベースを操作するのに対して，NoSQL は SQL を使

用しない，NoSQL は RDBMS の否定してできたものではなく，RDBMS が得意ではないことを実行可能にしたデータベースである．RDBMS と NoSQL の違いを以下に挙げる．

（1）データ構造
RDBMS では，データをテーブルという表形式で集約し，データ同士の関係性を定義する．そうすることで厳格なデータモデルを表現している．そのためテーブルのカラムを事前に定義しておく必要がある．一度定義したスキーマは固定的であり変更しにくい．NoSQL では，キーと対応するバリューの組み合わせ，あるいは，キー・バリューのペアと追加キー（カラムファミリー）によって表現されるため．データ構造がとても単純である．データ同士の関係も定義することができない．スキーマも定義する必要がなく柔軟に変更可能である．

（2）データの一貫性
RDBMS では，ACID（Atomicity Consistency Isolation Durability）特性が実装されているため，データの一貫性が厳密に維持されている．一方，NoSQL データベースでは ACID のような堅牢なものではなく，Eventual Consistency という実装になっており，最終的に一貫性が維持されるが，一時的には一貫性が厳密ではない．

（3）拡張性
RDBMS の場合，ACID やデータ構造を重視するため，データ量が増えたときはより大きなサーバに変えるスケールアップが基本であり，スケールアウトしにくいアーキテクチャとなっている．また，データの一貫性を厳密に行うためパフォーマンスの低下も起る．NoSQL データベースの場合，スケールアウトが容易にできる設計になっているため拡張性に優れている．また，パフォーマンスの低下も少ない．

（4）耐障害性
RDBMS はレプリケーションによって，複数のサーバにデータを複製し耐障害性を高めている．しかし，データの不整合が起った場合やレプリケーションを追加する際には運用上の負荷やコストが大きい．NoSQL の場合，分散環境で動作するため，単一障害が少ないものが多く，障害に対する対策コストが少なく済む．

以上のように，NoSQL は RDBMS の得意ではないこと解決する特徴が多い．

- クラウドコンピューティング
クラウドコンピューティングとは，ネットワーク，サーバ，ストレージ，アプリケーション，サービスなどの構成可能なコンピューティングリソースの共有プールに対して，便利かつオンデマンドにアクセスでき，最小の管理労力またはサービスプロバイダ間の相互動作によって迅速に提供され利用できるというモデルのひとつであると，アメリカ国立標準技術研究所は定義している．従来であれば，自分自身で物理

的なサーバを用意し，データを管理しなくてはならなかった，扱うデータ量が多くなればなるほど必要なリソースは多くかり，全てを管理するのはとても大変である．しかしクラウドコンピューティングの技術が発達したことによって，技術者は物理的なリソースやデータ管理について考える必要がなくなり，本来の問題であるサービスや研究に集中することが可能になった．また使用量は利用しただけ支払うという方式であり，資本の少ないスタートアップ企業や個人でも手軽に利用することができる．この敷居の低さが，ビッグデータ活用に大きく貢献している要因である．

2.4 ビッグデータの活用

ビッグデータは様々な場面で活用されている．商品やサービスのレコメンデーション，行動ターゲティング広告，異常検知，サービスの改善，渋滞の予測，風邪の予測，株価の予測など活用される対象は幅広い．ビッグデータの活用パターンは大きく分けて 2×2 の 4 つのパターンに類型化できる．軸は「個人最適/全体最適」「リアルタイム型/バッチ型」である．個人最適とは，分析結果の受益者が特定の個人やモノであり，特定の個人やモノにとって最適な情報やサービスを提供したり，最適な処理を促すものである．全体最適とは，分析結果の受益者が特定の個人やモノではなく，その個人やモノが属するコミュニティ，あるいは社会全体などマスの場合である．それぞれについて以下で説明する．

2.4.1 ビッグデータの活用パターン

ここでいくつか論文やサービスの例を挙げる予定！

- 個人最適・バッチ型
  個人最適・バッチ型とは，特定の個人やモノに関するデータを収集して，その人に最適な商品やサービスをレコメンドしたり，そのモノに最適な処置を行ったりする場合である．代表的なサービスが Amazon である．Amazon は多くのユーザの購入履歴，閲覧履歴を収集して嗜好の似ているユーザをグループ化し，嗜好の似ているユーザの情報から商品をレコメンドしている．そうすることで利用者に価値のある情報を提供し，購入を促進している．これは協調フィルタリングというレコメンデーションアルゴリズムである．

- 個人最適・リアルタイム型
  個人最適・リアルタイム型とは，特定の個人またはモノに関するデータを収集し，その人に最適な商品やサービスをレコメンドしたり，そのモノに最適な処置を行ったりする場合である．ただし，バッチ型とは異なり，その人に商品やサービスをレコメンドしたり，そのモノに最適な処置を施すタイミングはコンテキストに合わせてリアルタイムに行われる．

- 全体最適・バッチ型
  全体最適・バッチ型とは，多数の個人やモノが発信する情報を収集・蓄積し，蓄積したデータを一括して統計的に処理・分析することで，その個人やモノが属するコミュニティや社会全体にとって役立つ統計情報をフィードバックしたり，最適化を図る場合を指す．ただし，最適化したりするタイミングは問わない．

- 全体最適・リアルタイム型
  全体最適・リアルタイム型とは，多数の個人やモノが発信する情報を収集・蓄積し，蓄積したデータを一括して統計的に処理・分析することで，その個人やモノが属するコミュニティや社会全体にとって役立つ情報をコンテキストに合わせてリアルタイムにフィードバックしたり，最適化する場合である．

2.5 本論文の着眼点

第 2.1.1 節や第 2.1.2 節で述べた通り，電車には遅延，運転見合わせ，その対応に問題がある．そこで電車利用者が急な遅延や運転見合わせで混乱しないように，リアルタイムに電車のダイヤの乱れを知る方法が必要である．本研究では，Twitter ユーザを 1 つのセンサとしてみなし，発信される電車に関するテキストデータを第 2.3.2 節で説明した技術を用いて収集，蓄積，分析することによって，列車の運行状況をモニタリングし，ユーザに提供するアプリケーションを開発する．Twitter を活用する理由は，リアルタイム性を重視するためである．

2.6 まとめ

本章では，電車の遅延，運行見合わせ，その対応に問題があることを示した．またビッグデータの概念について述べ，ビッグデータを活用することによって，日常の様々な問題を解決した研究やサービスを挙げた．本研究では，Twitter に投稿された電車に関するツイートを収集，蓄積，分析することによって，列車の運行状況をモニタリングし，ユーザに通知するアプリケーションの開発を目指す．

第3章　関連研究

本章では，既存のデジタル情報を統合する既存研究について述べる．また，既存の情報統合に対する対策についても言及する．

3.1　ソーシャルネットワークを利用した情報収集

Krishnamurthy の論文 On the leakage of personally identifiable information via online social networks[5] で，ソーシャルネットワークサービス (SNS) を利用したプライバシの脅威について述べている．この論文ではユーザが SNS に登録する情報と他の情報を組み合わせる事で，個人が特定される危険性について述べている．例えば，二つの SNS を二つ以上組み合わせて，個人情報を複数取得し，ユーザのプロファイル作成を可能にする．SNS には，E-mail アドレス，住所に関する情報，本人の写真などを記載する場合があり，これらの情報を識別要素とすることで，個人情報を得る．

また，SNS から発行される Cookie を解析することで個人の識別要素が含まれていることを記している．Cookie には，直接ユーザの個人情報が含まれているわけではないが，ユーザ ID が含まれている場合がある．Cookie が外部のものでも利用できる Third-party Cookie である場合は，Cookie の情報と SNS の情報を照らし合わせることで本人を特定することができる．その攻撃モデルを図 3.1 に記す．

ここでは，Cookie とホストが送信する Request-URI によって，ユーザの Web 履歴と個人情報をマッチングする例を挙げている．Third-party Cookie の場合など個人情報がユーザの意図しないところで公開されていることや，SNS におけるユーザのプライバシの脅威についてに理解せずに，情報を書きこむことに対しての危険性を述べている．この論文で想定している攻撃手法は本論文で述べる第 2 章で述べるモデル図??と図??の組み合わせに該当する．

3.2　Web 上での情報収集

インターネット上での検索エンジンを利用することで，対象とするユーザの人間関係や，社会的な立場が明らかになる場合がある．Web 上の情報からの人間関係ネットワークの抽出 [6] では，検索エンジンを用いてターゲットとなるユーザの人間関係を抽出している．人間関係の抽出方法は，学会発表時の共著からユーザの人間関係を推測している．人間関係の分類としては，共著や発表，同研究室，プロジェクトの 4 つに分類している．これによって，ユーザの実社会における人間関係や研究分野などを知ることがで可能とな

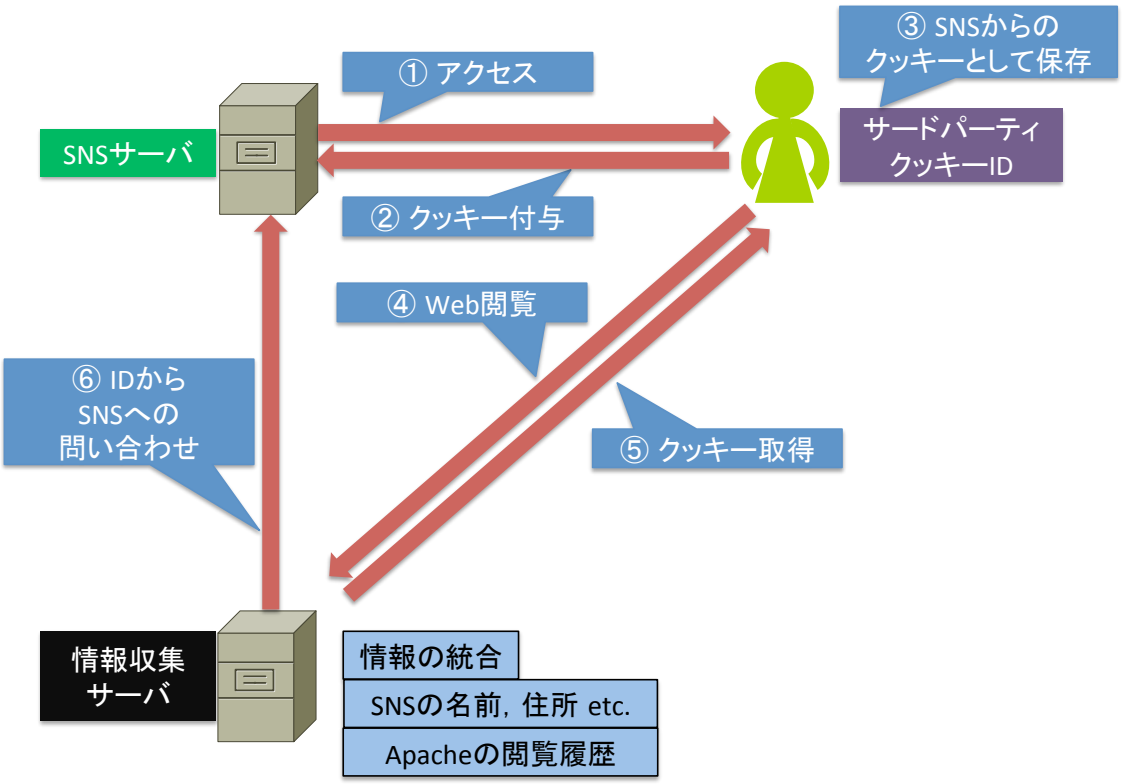


図 3.1: Cookie を利用して得た SNS 情報と Apache ログの組み合わせ手法

る．推測に利用している情報はすべて Web 上で公開している情報のみであるが，適合率は 8 割を超えるため非常に有用であると言える．このモデルは，ユーザが自ら Web ページを作成・公開するため，図??に該当する．

3.3 ベイズ統計を用いたユーザ嗜好の分析

事例ベース推論という研究とベイズ統計とよばれる統計研究を組み合わせることによってユーザの好みを検索する Profiling Case-Based Reasoning and Bayesian Networks[7] という研究がある．この研究はあらかじめデータベースに登録したデータを元にユーザの行動の頻度や傾向，他のユーザに対する影響度などを収集し分析することによってユーザを識別する．しかし，この研究は事前にユーザを登録する必要があり，取得する情報もデータベースが保有する情報しか利用できないという欠点がある．このモデルは，ユーザが自ら Web ページを作成・公開するため，図??に該当する．

3.4 ブラウザ情報を利用した個人識別

ブラウザの情報を利用することでユーザの識別が可能かというというプロジェクトがある [8]．この研究では User Agent string，プラグインのバージョンやフォントの設定などのデータを総合して，ホストやユーザを識別することは可能かを検証している．このプロジェクトではユーザのブラウザ情報を収集したデータベースをもとに，識別するプログラム公開することで，ユーザに，Web 閲覧などの情報を利用したトラッキングや広告に対する脅威を周知することを目的としている．このモデルは，ユーザが Web ページを閲覧することで情報を送信するため，図??に該当する．

3.5 情報統合に対する対策の検討

複数の情報を組み合わせることは昔から懸念されており，それに対する対策が検討されている．日本での事例をあげると，ネットワーク上での情報統合によるプライバシ侵害とその対策 [9] では，インターネットが今日よりも発展する前に，情報統合の対策が必要であるとして提案されている．この論文は日本の法律とドイツの法律を比較し，個人情報の組み合わせを守る仕組みを提案している．近年は，特に情報の組み合わせによる対策などプライバシ保護を視野に入れた手法を提案することが多くなっている [10][11]．また，情報の扱い方をはじめとしたユーザや開発者・管理者のガイドラインの提案を行っているところもある．個人情報・プライバシの保護 [12] では適切な情報の取り扱いやユーザのとるべき行動を示している．しかし，どのような情報がプライバシを明確にしていない．

3.6 まとめ

本章では，複数の情報を組み合わせることによって，ユーザのプロファイルを作成する手法について述べた．複数の情報を組み合わせることによって，単体の情報だけでは得られなかったユーザに関する情報を得ることができる．ユーザが同意を得て利用するサービスと別のサービスを利用して情報統合することで，プライバシの脅威となることを示している．このように，他にも個人情報を組み合わせ続けると，より正確な個人のプロファイルを作成できる．それとともに，情報統合に対する対策を考慮したシステムの例を挙げ，情報取り扱いのガイドラインを提示したが，どのような情報が組み合わせことが問題かを明確にされていない．したがって，どのような情報がプライバシを脅かすのかを明確にし，どのように取り扱うかのガイドラインを提示する必要がある．

第4章　提案手法

4.1　ネットワーク管理者と取得情報

4.1.1　前提

4.1.2　パケットのヘッダ情報

4.1.3　ホスト識別による調査

4.2　同一セグメント上のユーザと取得情報

4.2.1　前提

4.2.2　共有ホスト名

4.3　第三者であるユーザと取得情報

4.3.1　前提

4.3.2　Bluetooth

4.4　まとめ

第5章　実装

5.1　section1

5.2　section2

5.3　section3

第6章　評価

6.1　評価手法

6.2　内容ベース分類器部分の評価

6.3　時系列分析部分の評価

6.4　遅延発生から通知までの時間の評価

6.5　まとめ

第7章　結論

7.1　まとめ

7.2　今後の展望

謝辞

本論文の作成にあたり，ご指導頂いた慶應義塾大学環境情報学部学部長 村井 純博士，同学部教授 徳田 英幸博士，同学部教授 中村 修博士，同学部准教授 楠本 博之博士，同学部准教授 高汐 一紀博士，同学部准教授 三次 仁博士，同学部准教授 植原 啓介博士，同学部専任講師 重近 範行博士，同学部専任講師 中澤 仁博士，同学部専任講師 Rodney D.Van Meter III 博士，同学部教授 武田 圭史博士，同大学 DMC 機構専任講師 斉藤 賢爾博士，同大学政策・メディア研究科特別研究講師 佐藤 雅明博士に感謝致します．特に武田圭史博士は，研究で行き詰まる私に対して非常に根気強く指導していただきました．常に新しいアイディアと研究手法で私を導いていただき，何度も私に新しい視点や手本を見せていただきました．本当にありがとうございました．

そして，本研究を進めていく上で，様々な励ましと助言，お手伝いをいただきました，村井研究室卒業生である中村 友一氏，金井 瑛氏，奥村 祐介氏，海崎 良氏，石原 知洋氏，中里 恵氏，尾崎 隆亮氏，中島 智広氏に感謝致します．

慶應義塾大学大学院メディアデザイン研究科博士課程遠峰 隆史氏，同大学政策・メディア研究科後期博士課程 岡田 耕司氏，堀場 勝広氏，田崎 創氏，工藤 紀篤氏，久松 剛氏，松園 和久氏，三島 和宏氏，水谷 正慶氏，松谷 健史氏，空閑 洋平氏，同研究科修士課程，六田 佳祐氏，峯木 厳氏，江村 圭吾氏，黒宮 佑介氏，佐藤 龍氏に感謝致します．特に水谷 正慶氏は，博士論文の執筆や学会発表で多忙な身にも関わらず，親身に相談に乗っていただき，研究の方向性を指導や実装の細やかなケアをはじめとするあらゆる面で面倒を見ていただきました．氏なしでは卒論執筆だけでなく充実した研究室生活を送れませんでした．本当に感謝致します．

研究に協力をしていただいた，三部 剛義氏，中村 遼氏，福岡 英哲氏，中島 明日香氏，市川 博基氏，Doan Viet Tung 氏，鎌田 和大氏，梅田 昇翔氏，相見 眞男氏，中井 研氏，藤原 龍氏，吉原 大道氏，小澤 みゆき氏，澁田 拓也氏，村上 滋希氏と徳田・村井合同研究室の皆様，そして卒論執筆で迷惑をかけた DSAP09 メンバーに感謝致します．

研究室で苦楽を共にした永山 翔太氏，佐藤 貴彦氏，波多野 敏明氏，勝利 友香氏，朝永 愛子氏に感謝致します．彼らと一緒に研究をすることでお互いを刺激しあい，より質の高い議論や研究をすることができました．

私の大学 4 年間の心の拠り所であった SFC スペイン舞踊部と草本 麻里子氏をはじめとする部員全員に心から感謝致します．卒論執筆をする私を暖かく見守り続けてくれたデンスケと，常に場を和ませてくれた社長に感謝します．彼らのおかげで心に余裕をもって卒論執筆できたと確信しています．

最後に，大学入学からの 4 年間だけでなく 22 年間をあらゆる面で支えていただいた父，上原 健三，母，上原 昌子と私の家族に心から感謝致します．

参考文献

[1] 国土交通省. 旅客の輸送機関別輸送量・分担率の推移. http://www.mlit.go.jp/common/000232360.pdf 8月29日に閲覧.

[2] 東洋経済ONLINE. 満員電車も遅延も許せない！　通勤問題に特効薬はあるのか《鉄道進化論》. http://toyokeizai.net/articles/-/10756 8月30日に閲覧.

[3] JR東日本旅客鉄道株式会社. 列車運行情報サービス. http://traininfo.jreast.co.jp/train_info/kanto.aspx.

[4] Twitter. 小田急公式アカウント. https://twitter.com/odakyuline_info.

[5] B. Krishnamurthy and C.E. Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 7–12. ACM, 2009.

[6] 松尾 豊, 友部 博教, 橋田 浩一, 中島 秀之, and 石塚 満. Web 上の情報からの人間関係ネットワークの抽出. **人工知能学会論文誌** = *Transactions of the Japanese Society for Artificial Intelligence : AI*, 20:46–56, 20051101.

[7] Schiaffino Silvia N and Analia Amandi. User profiling case-based reasoning and bayesian networks. *7th Ibe-American Conference on Ai and Brazilian*, 2(1):19–22, 11 2000.

[8] Electronic Frontier Foundation. Panopticlick. http://panopticlick.eff.org/, 1 2010.

[9] 本村憲史 and 金田重郎. ネットワーク上での情報統合によるプライバシー侵害とその対策. **経営情報学会** *1998* **年春季全国研究発表大会**, *D-1-2*, pages 65–68, 1998.

[10] 佐藤 雅明. **インターネット上での自動車情報基盤の構築**. PhD thesis, 慶応義塾大学政策・メディア研究科, 2008.

[11] A. Tootoonchian, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr: Better privacy for social networks. *CoNEXT*, pages 169–180, 2009.

[12] 松井志菜子. 個人情報・プライバシーの保護. **長岡技術科学大学言論・人文科学論集**, 19:83–133, 2005.