卒業論文　　2013年度（平成25年度）

マイクロブログを活用した電車遅延の検出

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9 月25 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

本章では，背景であるユーザが意識せずに発信している情報が，ユーザプライバシを脅かす可能性があることを述べる．そして，新しいデジタル通信時代のプライバシのあり方を提案するという目的を明らかにするとともに，本論文の構成を記す．

1.1　ユーザが発信する情報とその利用

近年，情報技術の発展によってユーザに関する様々な情報が，デジタル通信上で送受信され，記録として残せるようになった．これによって，ユーザに関する情報を収集・解析することで，新しい価値を生み出すことができる．このため，よりユーザの要求に応じたサービスの提供が可能となった．代表的なものとしてはユーザの購買履歴と他ユーザの購買履歴を比較するレコメンデーション技術を利用している Amazon[1] や，現在の位置情報と広告を組み合わせることによって，ユーザの周囲の地理情報を得るサービス NAVITIME[2] などが挙げられる．しかし，このような情報技術の発展によって，これまでは個人情報と見なされなかった情報が，ユーザのプライバシを脅かすという懸念がある．

個人情報を利用するインターネットサービスやコンテンツが増加するにあたり，ユーザのプライバシを保護する必要性がある．そのため，情報保有者は厳密な管理が求められるようになった．個人情報を扱う側は，個人情報に関する規約を記載すると同時に利用プライバシを保護する様々な技術を提案している．本論文で述べる個人情報とは，個人情報保護法第二条一項に定義されている，”生存する個人に関する情報であって，当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む）”[3] を指す．そして，プライバシとは Privacy and Freedom[4] で述べている，”第三者が、自らに関する個人情報をどの程度取得あるいは共有することができるか、自ら決定できる権利”と定義する．

ユーザのプライバシが懸念される事例を数点挙げる．デジタルデバイスの増加によって，ユーザは意図せずに多くの情報を発信している場合がある．その際に，自身に関わる情報が含まれている場合や，その人と関わりがある情報を発信している場合がある．また，ネットワーク上での情報を複数を組み合わせることで，より正確に個人のプロファイルを作成することができる．これによって，複数の情報を組み合わせることによってネットワークにおけるユーザの調査や，調査によって得られた統計情報を公開することによる新しいサービスや，犯罪捜査などに利用できる反面，ユーザのプライバシが脅かされつつある．

現在，ユーザのプライバシを保護する多くの技術が提案されているが，デジタルデバイスの増加によって，ユーザが自身の情報を知らずに公開している場合や，今までは個人情報とならなかった情報を統合することによって，ユーザ自身のプライバシが脅かされる可能性がある．これは，どのような情報が自身のプライバシを脅かされているかのどの程度の権限を持つユーザまで知ることができるかという境界分けが曖昧だからである．この問題を解決するには，どのような情報によってユーザのプライバシが脅かされるのか明確にして，ユーザ自身もそれを知る必要がある．

一方，コンテンツやサービス提供者は個人情報をより効率的に取得する必要がある反面，同時にユーザのプライバシの保護も考慮に入れなければならない．取得することができる情報すべてを利用した場合，ユーザのプライバシの脅威につながる可能性がある．そのため，コンテンツ，サービス提供者は取得するユーザの情報を明記するか，もしくは制限をした上で情報を収集する必要がある．

1.2 本研究の目的

本論文の目的は，デジタル情報時代における新しいプライバシのあり方を提示することである．どのような情報がユーザのプライバシを侵害するかを明確に示すとともに，情報の取り扱いについて検討する．多くのデジタル機器を日常的に利用するようになったことで，ユーザもサービス提供者もどのような情報が発信されており，どのような影響をもたらすか把握しきれていない．そこで，本論文よって得られた知見を提示することによって，ISP をはじめとするネットワーク管理者や，サービス提供者，アプリケーション開発者は，どのような情報がユーザのプライバシを脅かすかを把握し，個人情報を明確な指針のもとに保護することができる．また，ユーザ自身も，どんな情報を守らなければならないかを知ることで，自身のプライバシを守らなければならない．プライバシの保護と，利便性はトレードオフであり，どこまで許容されるのか明確な区分けが必要である．ユーザのプライバシが容易に侵害されない社会を実現するためには，ネットワーク管理者，サービス提供者とユーザ側の双方から個人情報に関する保護をしなければならない．特に，ユーザが自身に関する情報を管理することが，ユーザのプライバシに対する脅威を低減できる可能性が高い．ユーザが利用する情報を管理することによって，情報の取捨選択ができるからである．しかし，これはユーザがどの情報を発信してよいか知っていることが前提である．

本論文ではユーザが無意識に発信している情報によって，ユーザのプライバシがどの程度脅かされているかを調査，実証する．この取り組みによって，ユーザ自身がどのような情報を他のユーザに取得されるとプライバシを侵害される可能性があるかをまとめるとともに，どの情報がどの立場のユーザまで知ることができるかという区分けを明確にする．例えば，ユーザの行動履歴はネットワーク管理者だけでなく同じネットワークに接続しているユーザまで知り得るかという調査などが挙げられる．

本論文では日常的にユーザが発信する情報を利用することで，プライバシが侵害される可能性のある手法を 3 つ示し，に対して検証・考察する．これによって，どこまで個人

情報を取得することができるのかを明確する．本研究はパケットのヘッダ情報，共有ホスト名，Bluetooth デバイス名を用いた場合に，個人情報を取得する手法を提示し，検証する．これらの情報は，ユーザが自ら発している情報であるため，容易に取得可能である．そのため，これらの情報がユーザのプライバシを脅かしているか確かめる必要がある．

これらの検証結果を踏まえて，個人情報利用時におけるプライバシを保護を目的としたガイドラインを提案する．これにより，ユーザ自身やネットワーク管理者，アプリケーション開発者などの立場別に，プライバシ守るためには，何の情報を守り，または発信しても問題がないかという線引きを明確にすることを目的とする．

1.3 本論文の構成

本論文は全 7 章から構成される．第 2 章では，デジタル情報の収集方法とユーザのプライバシの脅威となる情報について述べる．第 3 章では，第 2 章で述べた課題に取り組む関連研究を紹介する．第 4 章では，どのような情報を組み合わせるとユーザの脅威となるかを調査し，その手法を提案する．第 5 章では，第 4 章で述べた手法を検討し，その実現結果について述べるとともに，考察を行う．第 6 章では，第 5 章で検証した手法の対策として，ガイドラインを提案する．最後に第 7 章で本論文の結論と，今後の展望を述べる．

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[5] 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 [6] 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

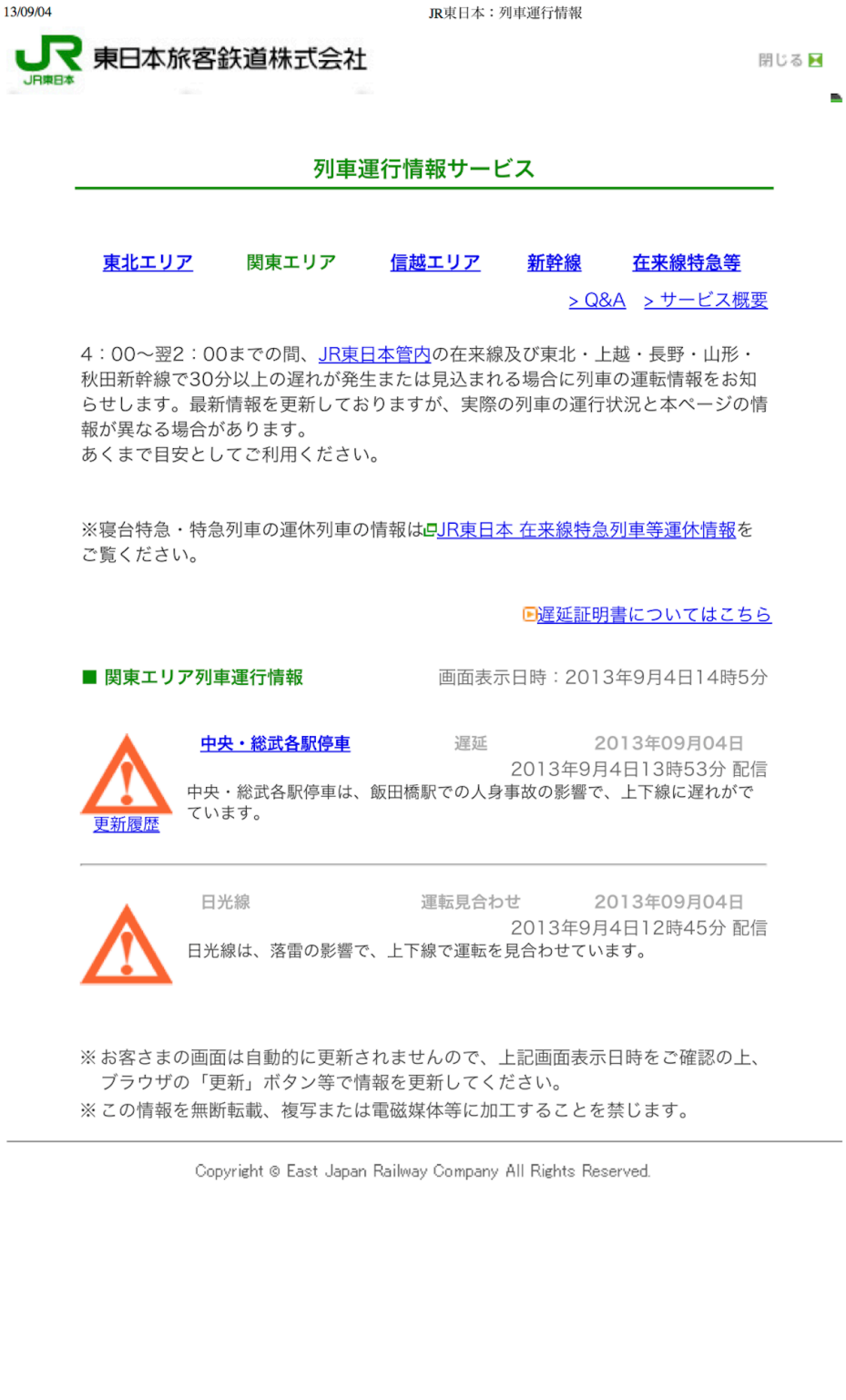
応としては，電工掲示板に状況の表示や駅で遅延証明書の配布などが行われている．リアルな対応はその場に行かなければ電車の運行状況を知ることができないため，利用者にとってはとても都合が悪い．そのため，駅に行かずに電車の運行状況を確認できるように各鉄道会社はオンラインで情報を発信している．オンラインでの情報発信の仕方を以下に挙げる．

(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[7] 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[8] 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，センサーデータ，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ーや放射線を測定するセンサーなどのセンサーデータ，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

2.4 ビッグデータの活用

2.5 本論文の着眼点

本論文は，第??節で述べた，ユーザが発信する情報を受動的に取得できる手法に着目し，中でも第??節で述べたホスト情報を対象とする．ホストが発信する情報はユーザのプライバシとなる要因を多く含んでいるため，ホストの情報のプロファイルを作成することはユーザのプライバシを脅かす可能性が高いためである．また，ホストが定常的に情報を発信しているため，情報の取得が容易であることも一因である．これらの理由から，本論文ではホストやユーザのプロファイルを作成する手法を提案する．これらの情報を利用することによって，ユーザが常に発信している情報がプライバシを脅かす可能性があるかを検討する．

2.6 まとめ

本章では，デジタル情報の収集を 3 つのモデルに分類し，その中でユーザの識別子となる情報の具体的な例を挙げた．ユーザのプライバシは法律や技術で保護されているが，複数の情報を組み合わせることによって，脅かされる場合があることを示した．しかし，サービス・コンテンツ事業者はユーザから取得した情報から，ユーザに応じたサービスを

提供する必要がある．そのため，ユーザは個人情報を提供するかわりに，サービスを受けるというトレードオフが少なからず存在する．そこで，それに伴った個人情報を考慮した情報技術が必要である．また，ユーザは自身に関する情報をどの程度発信しているか知り，何を守るべきかを明確にする必要がある．そして，どのような情報を組み合わせるとプライバシの侵害になるかを検証する．

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[9] 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 [10] 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

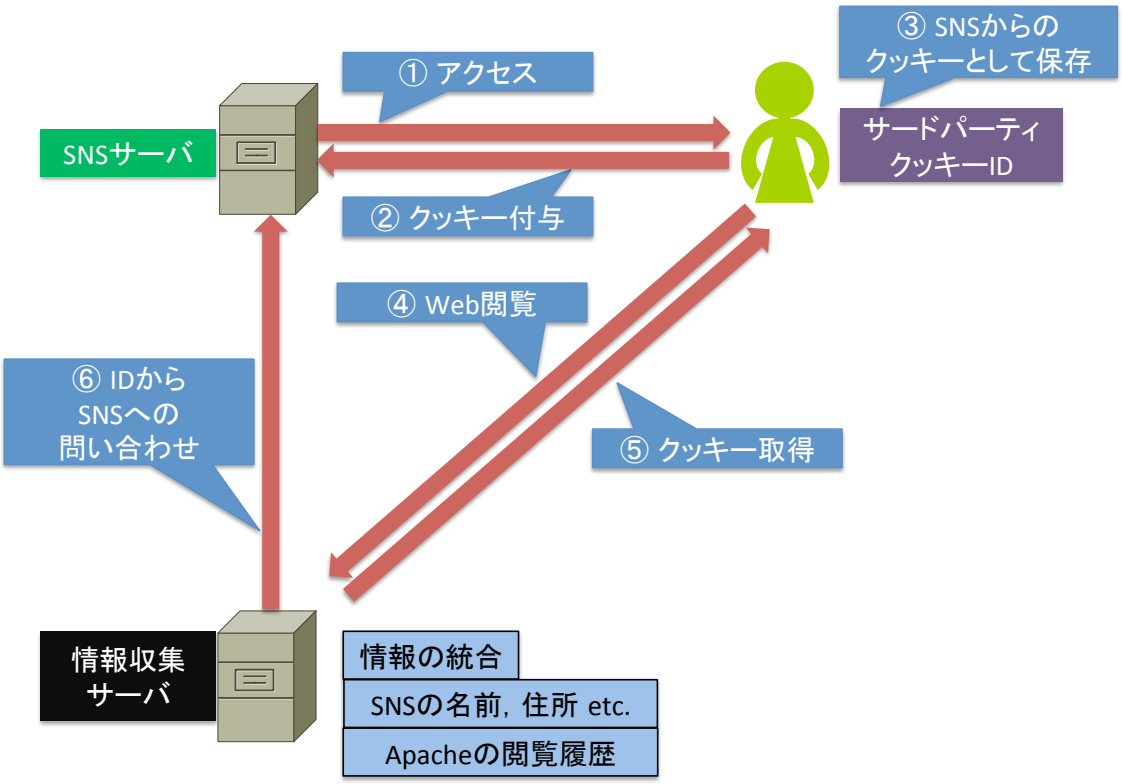


図 3.1: Cookie を利用して得た SNS 情報と Apache ログの組み合わせ手法

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[11] 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 [12]．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 [13] 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 [14][15]．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 [16] 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　デジタル情報を用いたユーザ特定手法

サービス・コンテンツ提供者は，ユーザに応じた効率的なサービスを提供することが求められる．様々なユーザのプライバシを考慮しながらも，ユーザの情報を収集する必要があるので，ユーザの同意を得るなど制限を付けて，取得する情報としない情報を明確にする必要がある．

第2章で述べたように，情報収集者が取得する情報は，対象ユーザとのネットワーク上の関係によって変化する．そのため，情報収集者と対象ユーザのネットワーク上の位置関係ごとにユーザのプライバシに影響を与える手法を提案した．図4.1に観測者と利用情報ごとに分けた手法を示す．まず，ネットワーク管理者はパケットのヘッダ情報を利用したホストの特定をする．次に，同一セグメントのユーザは，サービス探索情報を利用したプロファイルを作成する．最後に，同一ネットワークに接続していないユーザは，Bluetoothを利用したプロファイルを作成する．これら3つの手法について述べる．

4.1　ネットワーク管理者と取得情報

ネットワーク管理者にとって収集が容易であるものの一つに，トラフィックデータが挙げられる．しかし，ユーザを一意に識別できる，パケットのペイロードやMACアドレスなどはネットワークのポリシや構成によって取得できない場合がある．そこで，パケットのヘッダ情報に焦点を当て，パケットのヘッダ情報のみから，ホストを特定する手法を提案する．

4.1.1　前提

パケットヘッダ情報取得によるプライバシの脅威に関する手法の前提について述べる．パケットの情報を取得するネットワーク管理者は，ネットワークのポリシを自由に設定できるISPやネットワーク管理者を想定する．図4.2に示すように，管理者がパケットのヘッダ情報を収集する機器を，ネットワークの中継地点で設置することによって，ネットワークトラフィックを取得する．ネットワーク構成は一箇所のみ外に出る回線が存在し，同じネットワークにおいてユーザの発信するパケットは必ずその機器を通過するものとする．

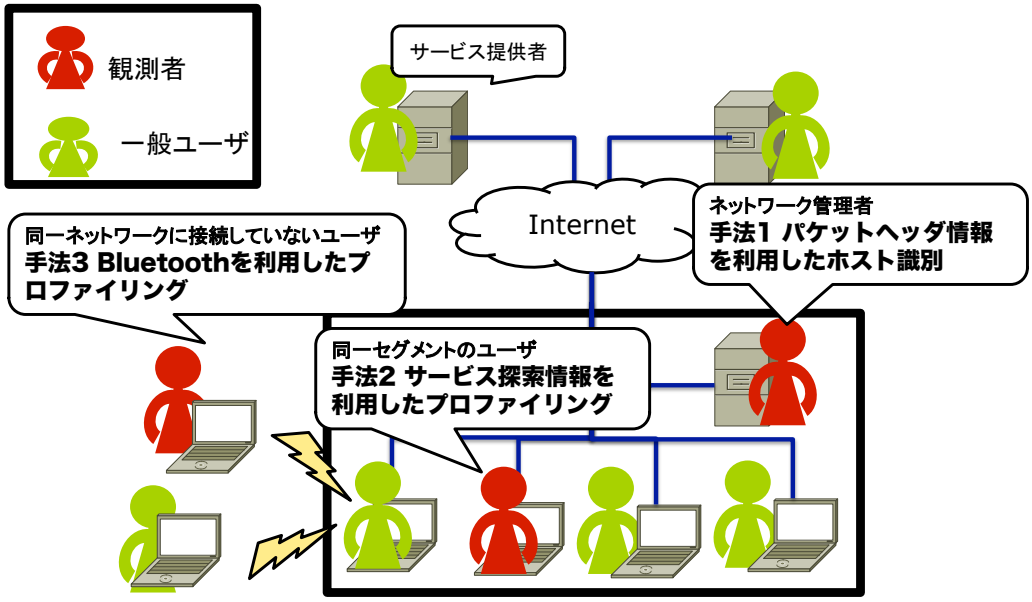


図 4.1: ユーザ特定手法の全体図

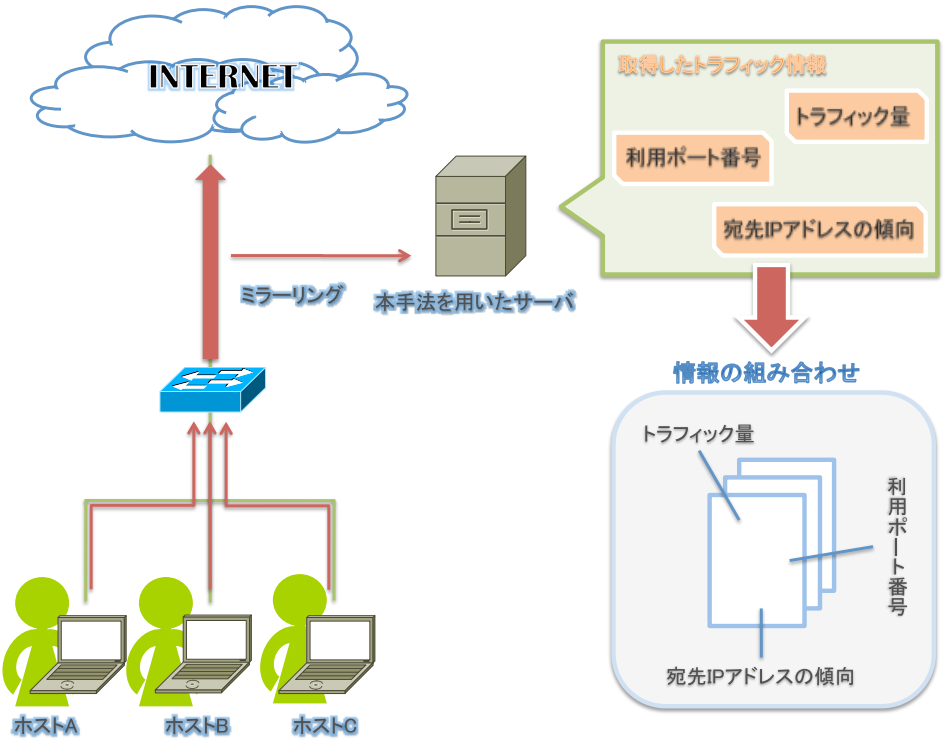


図 4.2: パケットヘッダ情報の収集システムの概要

4.1.2 パケットのヘッダ情報

パケットのヘッダ情報はネットワーク上を流れるトラフィックを観測するだけで取得することができるが，ペイロードを含む情報を得る場合，ユーザの個人情報と密接な関わりがあるためユーザの同意が必要である．しかし，ネットワーク管理者が，ネットワーク帯域の制御，ネットワーク上の通信の統計やトレンドを利用したい場合，パケットを閲覧するために全てのユーザの同意を得ることは困難である．そのため，ペイロードは閲覧せずにヘッダ情報のみでネットワーク全体のトラフィックを収集する手法が提案されている [17]．ここで述べているパケットのヘッダ情報は送信先，発信元 IP アドレス，ポート番号，プロトコルを指す．しかし，パケットのヘッダ情報を収集する手法が，ユーザのプライバシを脅かさないとは断言できない．

各ユーザはそれぞれの利用状況に応じた特徴的なパケットを送受信している．ヘッダ情報からは，送信先・発信元 IP アドレス，ポート番号から使用しているサービス，アプリケーションの使用頻度の情報が取得できる．そして，送信先 IP アドレスからはユーザの通信相手の情報が把握可能である．これによって，該当ユーザはどのようなホストと通信する傾向があるかを把握できる．他にもヘッダ情報のみで OS を推測する Passive Finger printing[18] を利用することで，ホストの識別要素となる．このように，多くの技術を組み合わせることによって，ホストに関する情報を蓄積できる．この蓄積した情報をもとにホストの特定をする手法を提案する．接続頻度が多いホストや，起動時の挙動，プロトコルと転送量を収集することで，ホストがアクセスする傾向のあるサイトや，利用しているアプリケーション，ネットワーク上での挙動を推測できる．

4.1.3 ホスト識別による調査

実際にパケットのヘッダに含まれる情報が識別要素として成り立つのかという事前調査調査を行った．以下に個人の特定に利用できる情報を述べる．

個人の特定に利用できる情報

- **送信先 IP アドレスとポート番号の組み合わせ**

  送信ポート番号と IP アドレスの関係に着目して調査を行った．その結果，識別要素として想定していた IP ポート番号による識別は困難であることが判明した．図 4.3 にホスト A の送信先 IP アドレス・ポート番号，図 4.4 にユーザ B の送信先 IP アドレス・ポートを記す．X 軸が IP アドレス，Y 軸がポート番号を示す．

  二つの図からホスト A，B の明確な差異を見つけることができなかった．しかし，SSH や IMAP の利用先など，プロトコルと接続先 IP アドレスによっては，ホストごとに特徴を得ることができた．このことから，送信先 IP アドレス，ポート番号の組み合わせはポート番号によってはユーザの識別要素としては利用できるという結論に至った．ホスト識別に利用できるプロトコルは，IMAP，SSH，VPN などが挙げられる．

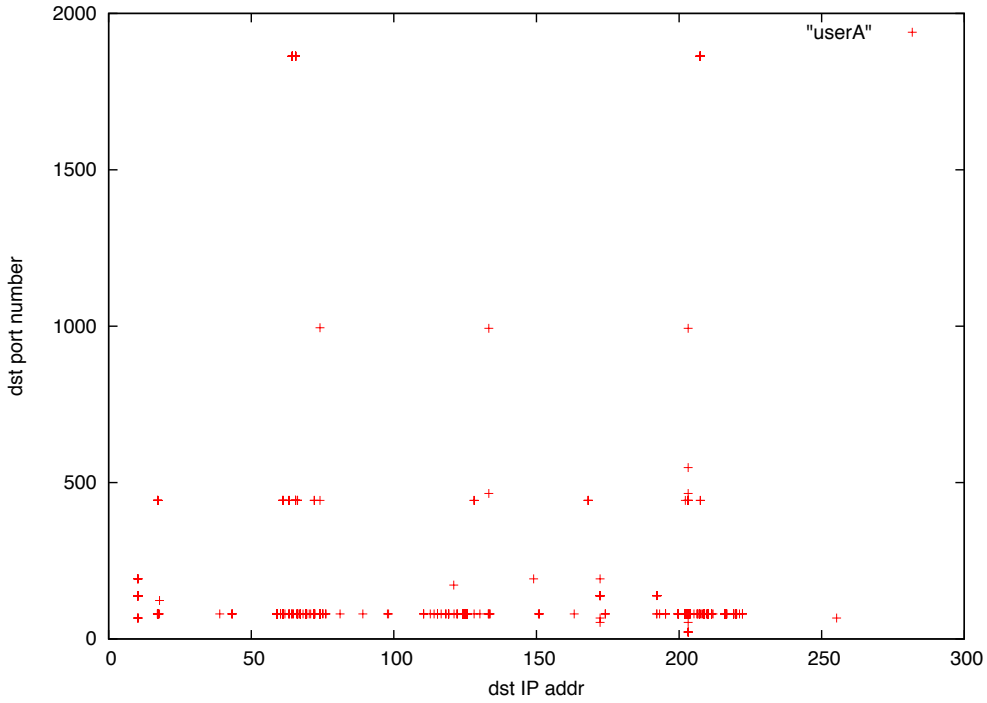


図 4.3: ホスト A の送信先 IP アドレス・ポート番号

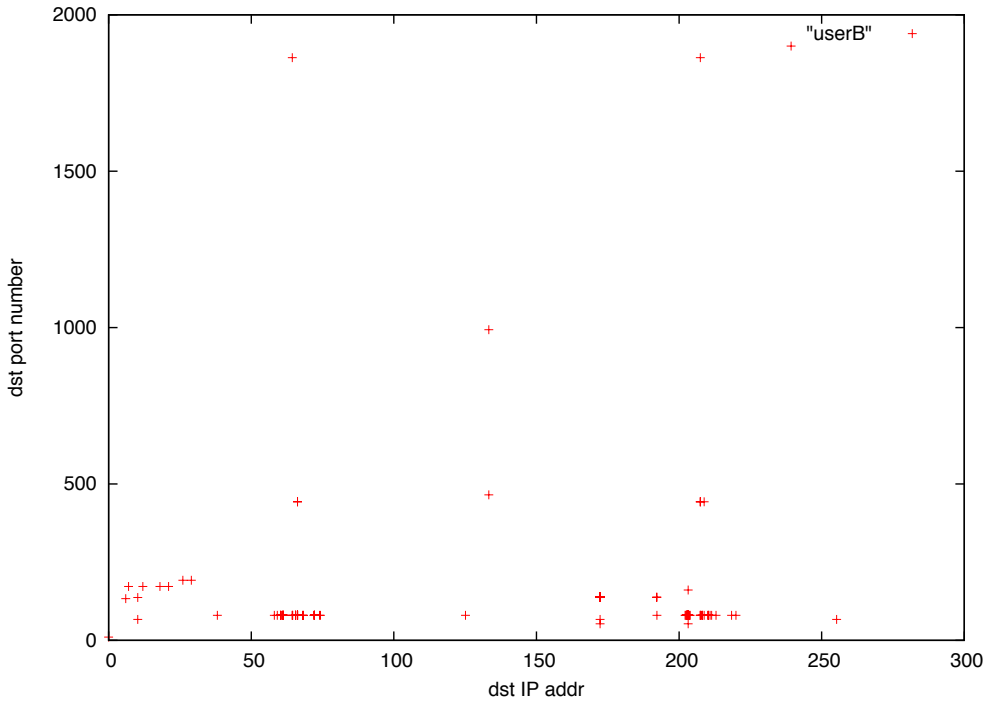


図 4.4: ユーザ B の送信先 IP アドレス・ポート番号

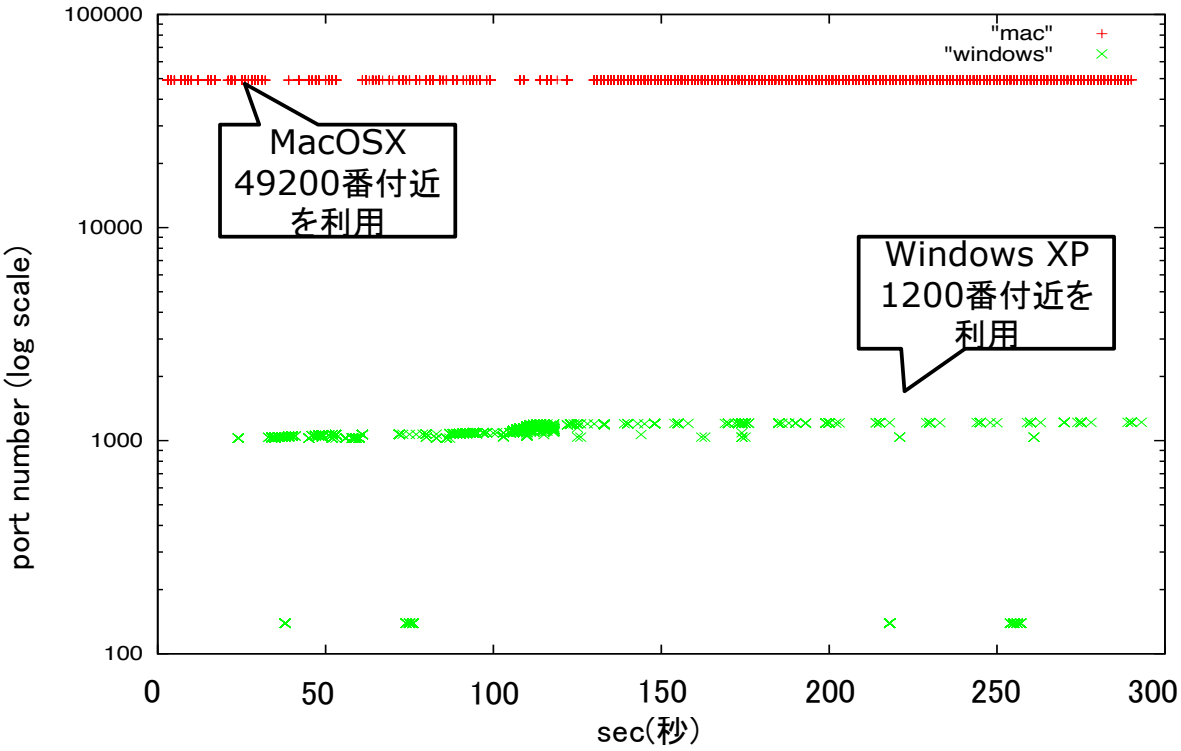


図 4.5: MacOSX と WindowsXP の利用送信元ポート番号

- **発信元ポート番号**

  OS ごとに発信元ポート番号は異なり，特徴がある可能性がある．そのため，発信元ポート番号の利用を比較する．その結果を図 4.5 に示す．図 4.5 の MacOSX と Windows XP の発信元ポート番号を比較した場合，容易に差を発見することができる．Windows XP と MacOSX をさらに長期的に観測したが，同じソースポート番号を利用されることはなかった．これは，OS ごとに送信元の利用ポートの傾向が違うためである．つまり，送信元ポート番号を収集することは OS を推測できる．他 OS を計測した場合，発信元ポート番号を利用した検証結果は表 4.1 にまとめる．
  以上のことから，IP アドレスと送信元ポート番号は識別要素として利用することはできないが，送信元ポート番号を得ることで OS を分類することができる．

- **パケットの発信タイミング**

  ホストの識別にあたり，最初の起動時間から数分の間パケットのヘッダ情報を観測することでホストを個別に特定することができるという仮説を立てた．実際に，ホストは OS やスタートアップに登録してあるアプリケーションが起動時に立ち上がるため，ユーザのパケットの発信タイミングからホストを識別できる可能性がある．ホストによっては，利用しているアプリケーションやサービスなど様々な特徴があるため，各ホストが発信するパケットのタイミングを識別要素として利用できる仮定した．そこで，OS が起動してからユーザが操作をするまでのタイミングに焦

表 4.1: OS と利用発信元ポート

OS	発信元ポート
FreeBSD	49152-65535
Ubuntu	32768-61000
Windows XP	1024-5000
Windows Vista	49152-65535
MacOSX	49152-65535

点を当て，個人差が出るかの調査をした．タイミングを収集するとして，ホストに IP アドレスが付与されてから約 2 分間の情報を調査する．しかし，パケットの取得タイミングはネットワークによって依存するため，同じネットワーク上で取得するという条件のもとパケットのトラフィック取得を行った．タイミング取得において，どの要素が毎回確認される共通のデータであるのかを取得した．実際に 7 回 OS を繰り返して再起動を行い，IP アドレスが付与される最初の 2 分間のタイミングを取得した．

今回は MacOSX version 1.6 と Windows Vista の OS を対象とし，クリーンインストールした状態で検証を行った．まず，MacOSX の結果を図 4.6 に示す．X 軸が IP アドレスが付与されてからの経過時間であり，Y 軸が再起動回数である．グラフの点は，ホストがパケットを発信した際に描かれる．次に，Windows Vista の場合のタイミングを図 4.7 に示す．

MacOSX に関しては，最初の数分間において一定の通信を繰り返す傾向が強く，比較的に特徴の出やすい結果であることが分かる．特に，起動してから 1 秒後から 2 秒後が特徴的であり，このタイミングを取得することで識別要素として利用できる可能性があると推測できる．それに対して Windows の場合は IP アドレスが付与されてから数十秒は連続的な通信が多いため，Windows と MacOSX と大きな違いが出たと言える．特に顕著なのが MAC OS と同じく 1 秒から 2 秒の挙動である．この挙動の間隔の差を利用することによって，OS の特定が可能であると言える．

次に，各パケットのプロトコルとポート番号を示す．二つの OS のパケット送信タイミングを種類に分けて更に詳しく解析したグラフが図 4.8 と図 4.9 である．プロトコルごとに分類し，ポート番号別に表示した．図 4.8，図 4.9 は X 軸が IP アドレスが付与されてからの時間に対して，Y 軸はポート番号を指している．また，色によって UDP，TCP，ICMP の 3 種類によって分類している．これによって，mDNS や SMB，DNS などの通信が多く観測できる．ここからパケットの共通項を取得する．他にも，毎回挙動が違うパケットを排除する．例えば，LDAP や NetBIOS など，ネットワークを構成するホスト群に作用されるパケットを排除する．

今回の前提として，ネットワーク管理者を想定しているため，MSND や NetBIOS などのマルチキャストやブロードキャストの通信を取得できない可能性がある．そ

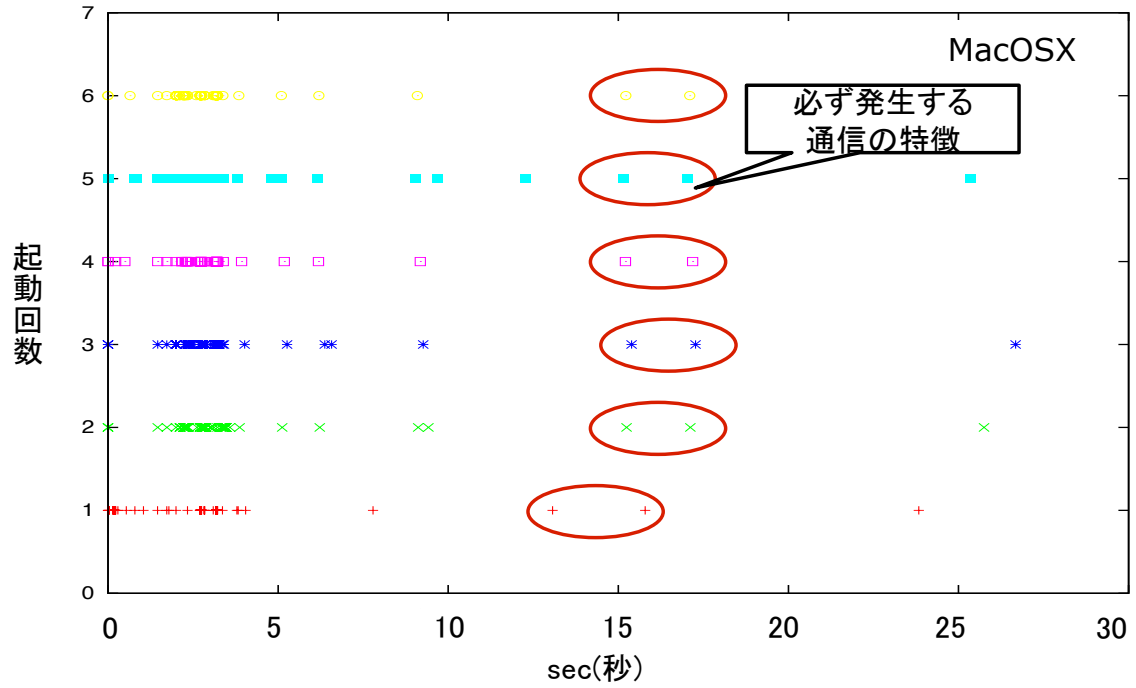


図 4.6: MacOSX の起動時のパケットの発信タイミング

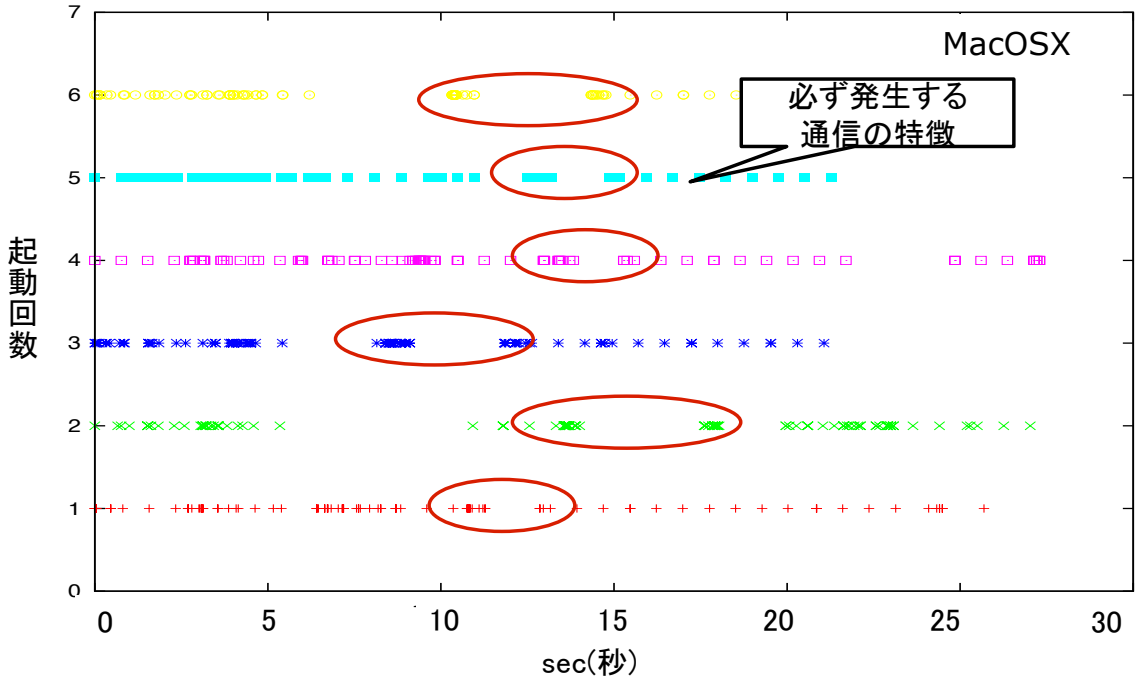


図 4.7: Windows Vista の起動時のパケットの発信タイミング

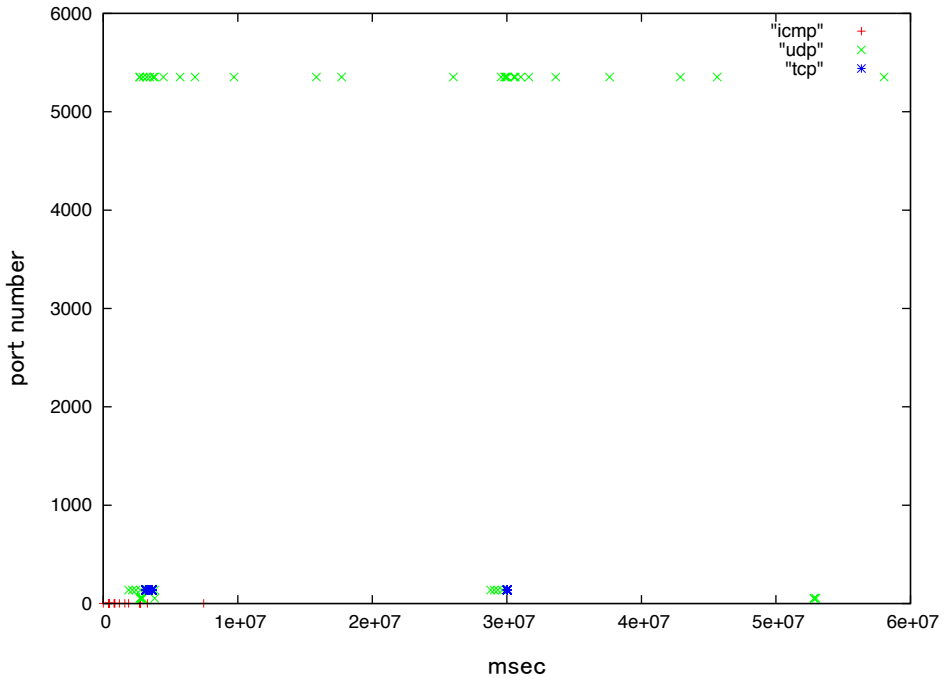


図 4.8: MacOSX のポート番号とプロトコル別パケットの発信タイミング

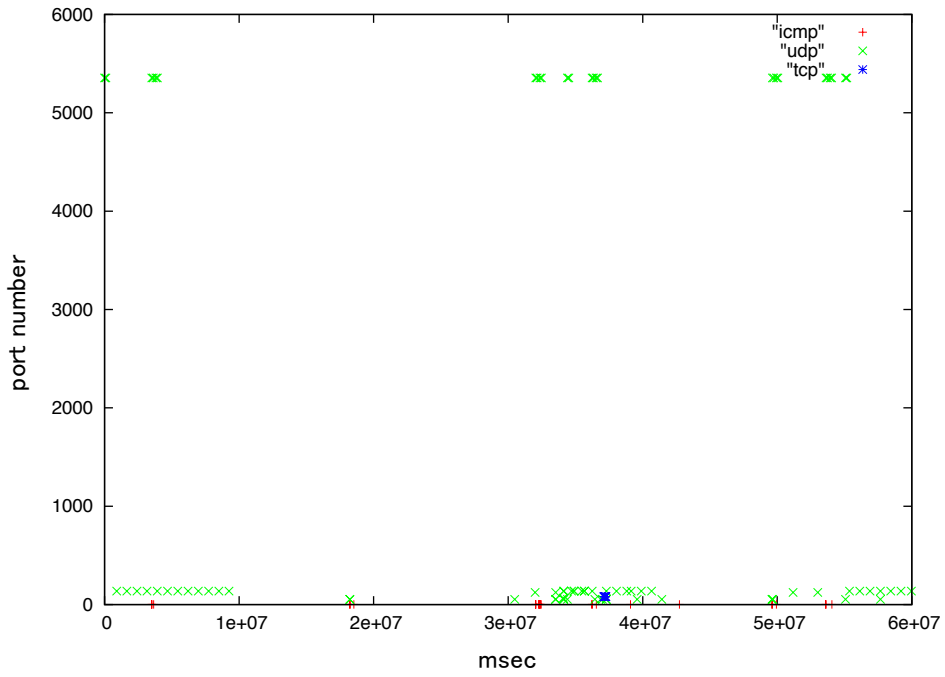


図 4.9: Windows Vista のポート番号とプロトコル別パケットの発信タイミング

こで，マルチキャストやブロードキャストを除いた場合は，これらの情報を取得することが困難である．特に，MacOSX で容易に発見できた図 4.6 でのホストの特徴はマルチキャストであるため，タイミングのみでホストを識別することのは困難である．MacOSX と同じく mDNS や SMB のパケットが多く送信されているが，Microsoft 社のサーバに対して HTTP 通信をしているため，TCP のパケットを発信するタイミングに着目することで容易に発見できる．

そこで，パケットの発信のタイミングを起動してからの数十秒後に焦点を当てる．実際にアプリケーションが起動する時間を計測したところ，IP アドレスが付与されてから約 60 秒後に事前に登録してあるアプリケーションが起動していたため，60 秒後の挙動に着目した．60 秒から 120 秒の間にパケットの発信があった場合，なんらかのアプリケーションが自動起動に設定されている可能性がある．図 4.8 にある図の TCP は NetBIOS セッションサービスであり，識別子としては除外される．このパケットの送信タイミングとプロトコルを利用して，自動起動に設定されていないホストと設定しているホストに分類できる．

次に，自動起動に設定しているアプリケーションを特定する必要があるが，非常に多くのアプリケーションがあるためすべてを網羅することは不可能である．そこで，今回は利用者が比較的多い IMAP と MSN メッセンジャを対象として取り上げる．

IMAP や MSN メッセンジャは利用するポートが固定されており，imap は 993 番であり MSN は 1864 番である．このポート番号から取得するパケットのトラフィック量から個人の識別要素として利用できる可能性が挙げられる．例えば，MSN メッセンジャなどは連絡先のユーザのリストを保持しているため，そのリストをダウンロードしなければならない．それを利用することによって，どの程度リストを登録しているのかを推測できる．同様に，IMAP を利用するためにはサーバからフォルダやメールのリストをダウンロードする必要がある．このトラフィック量からユーザの識別要素として利用することができる．

上記の方法はホスト起動時の数分間の挙動をもとに推測を行っているため，ネットワークに参加するユーザが電源を落とした状態から起動したか，サスペンドもしくはハイパネーションから復帰したかによって，取得するべき情報や除外すべき情報は大きく変化する．その起動か休止状態からの復帰かを判断する方法にパケットの発信タイミングを識別要素として利用することができる．これらの手法を用いて，IP アドレスが付与された時間からパケット発信のタイミングを計測することで，ネットワークに接続したホストが，起動したのか，復旧したのかを容易に判別することができる．図 4.10 は IP アドレスが付与されてからのホストの起動時と復旧時のパケットの送信タイミングの比較である．X 軸が IP アドレスが付与されてからの時間に対して，Y 軸は起動時の挙動，復旧時の挙動といったラベルである．

同じクリーンインストールを利用した MacOSX であるのに対して，明確な差が生まれた．復旧時は電源が落ちている状態からの起動時に比べて，早くホストを利用することができる．サスペンドもしくはハイパネーションはアプリケーションや OS に必要な機能を立ち上げている状態で行われるためある．また，最初の数十秒の挙

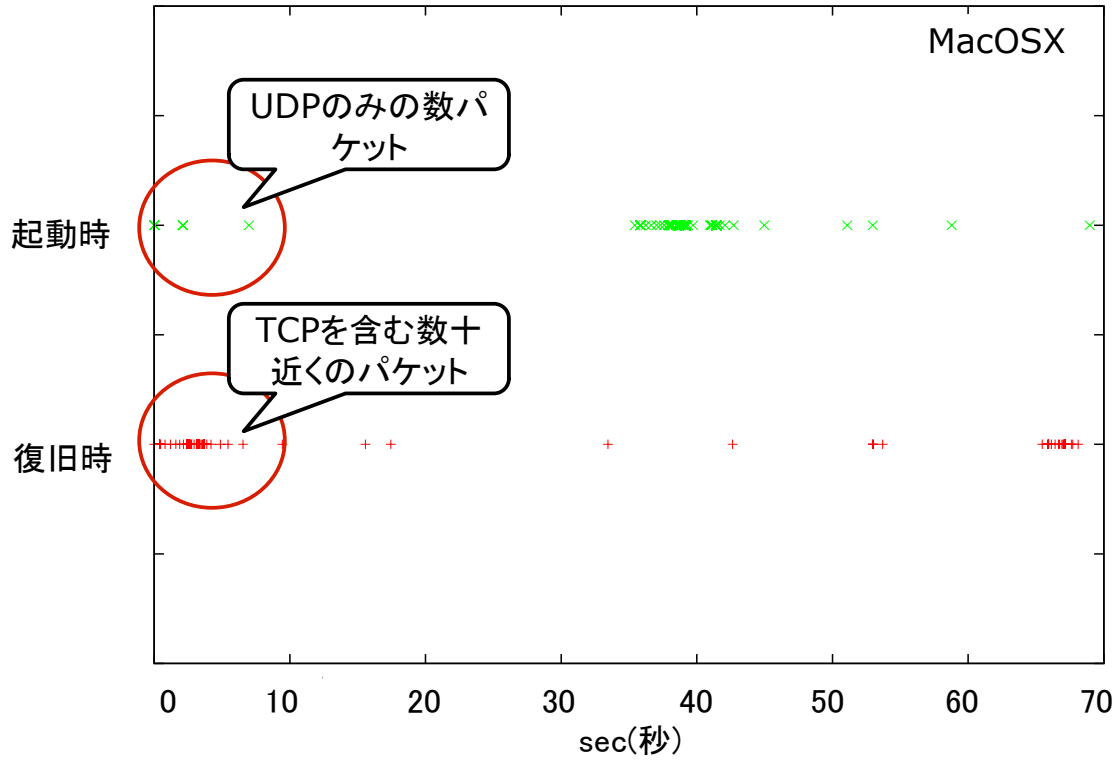


図 4.10: 起動時と復旧時におけるパケットの発信タイミングの比較

動は同じ OS であればどのホストでも同じであるため，起動して数秒で異なる挙動をした場合，休止状態から復旧したと判断することができる．

OS が休止状態から復旧した場合は，個人特定の重要な識別要素となる．OS が利用する発信元ポート番号は異なることは前述した通りである．休止状態から復旧した場合，発信元ポートは連続したポート番号を利用する．つまり，休止前に利用されていた発信元ポートの数個先のポート番号を利用する傾向がある．そこでホストの利用している最後の発信元ポート番号を保存することによって，そのホストが同じネットワーク上で復旧した場合，ホストの識別要素となり得る．しかし，一度ホストを再起動すると発信元ポートは初期値に戻るため，送信元ポート番号を利用した識別手法は復旧したホストにのみに利用することができる．

- **ホストが利用するサービス**

  ユーザが利用する送信先アドレスや，ポート番号から分かるサービスを識別要素とする．例えば，Web サービスを使用する場合，利用している Web サイトによってユーザをプロファイリングできる．サービスの利用頻度や傾向はユーザごとに差異あるため，ユーザの識別要素としても有効である．このため，Web アプリケーションや SNS を利用する頻度や時間帯を各ユーザごとに調べることによって類似するユーザを調査する．例えば，mixi[19] や Twitter[20] などにアクセスする時間帯や間隔を各ユーザごとに記録する．これらはユーザ特有の傾向であるためユーザ識別の

表 4.2: 識別要素とする対象ツール一覧

対象ツール	主な機能
Thunderbird	メールクライアント
Outlook	メールクライアント
firefox	Web ブラウザ
safari	Web ブラウザ
Omunigraffile	描画ツール
Skim	pdf 編集ツール
Windows update	OS アップデート
MacOSX software update	OS アップデート
GOM player	メディアプレイヤ
Quick Time	メディアプレイヤ
VLC media player	メディアプレイヤ
Windows media player	メディアプレイヤ

識別要素となり得る．また，ユーザの所属するネットワークの Mail サーバ，Web サーバへのアクセス情報も重要である．そこで本研究は各ユーザのアプリケーションやサービスの利用頻度やアクセスする間隔を用いてユーザを識別する．

- **利用アプリケーション**
  ホストの利用するアプリケーションによってユーザのプロファイル作成は可能である．表 4.2 にホスト識別要素とする対象アプリケーションを示す．次に，Web アクセスなど履歴からユーザの興味動向を推測する．HTTP プロトコルと IP アドレスの接続頻度を利用できる．また，OS 起動時に自動的にアプリケーションが同時に起動する場合がある．その際に，アクセスする先を記録することでホストを分類する識別要素となる．これらの識別要素を利用して，ホストの分類，最終的にはプロファイルを作成する．

- **パケットのオーダ**
  パケットが発信される順序を取得し，IP ヘッダや TCP ヘッダを組み合わせて解析することによって，ユーザの OS 情報を知ることができる．Passive OS Fingerprinting を利用することによってユーザの所有しているホストを把握する．

以上がパケットのヘッダ情報を利用してホスト識別する手法に採用した情報である．

個人の特定に利用が困難な情報

- **DHCP から割り当てられる IP アドレス**
  ネットワークにおける通信では IP アドレスはホストの識別要素として利用されて

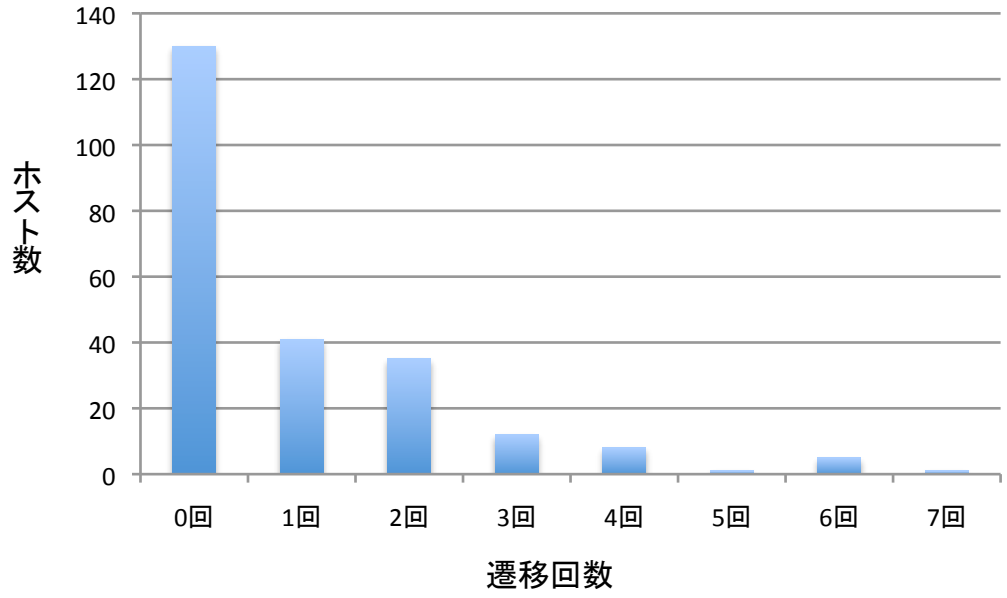


図 4.11: 時間におけるホストの IP アドレス遷移回数

いる．IP アドレス単体に焦点を当てた場合，ホストやユーザとしての識別要素として成り立つのかが問題がある．そこで，DHCP サーバのログを 2009 年 7 月 12 日 6 時から 7 月 18 日 6 時まで取得し，分析した．図 4.11 は筆者が所属する研究室のネットワークにおけるホストの IP アドレス遷移を示す．取得した IP アドレス数は 233 に対して取得した MAC アドレス数は 321 であった．そのうち，同一と見られるホストの IP アドレス遷移は 101 回あり，期間中に 7 回も IP アドレスがつけ変わるホストが存在した．つまり，観測した 6 日間に，43%のホストの IP アドレスが変化する結果になった．DHCP サーバの設定上，過去に IP アドレスを付与したホストには同じアドレスを割り振られる．また，同じホストが継続してネットワーク参加する場合も同じ IP アドレスが割り振られる．図 4.11 でも挙げられるように，一度も IP アドレスが変化しなかったホストは 130 台あった．しかし，付与する IP アドレスの範囲以上にホストが存在した場合，IP アドレスはつけ変わってしまう．そのため，ネットワークにおいてホストやユーザの識別に，IP アドレス単体では識別要素として利用できない．

- **起動時間・接続頻度**

  ユーザがホストをネットワークに接続した時間や接続時間帯の規則性を記録して，本人の生活習慣をユーザの識別要素とする．人間の生活習慣は多少のぶれが生じるが，傾向を把握することによって，パターンを取得できる可能性がある．そのため，ユーザがネットワークに接続する頻度やその接続時間はユーザの識別の材料となる．これに加え，本システムはホストを起動時に一番最初に利用するプロトコルや通信傾向も識別要素とする．長期的にネットワークトラフィックを取得した場合ユーザ

表 4.3: 送信先 IP アドレス上位リストの類似調査

	ホスト A	ホスト B	ホスト C
1 回	1	0	0
2 回	0	1	0

ごとの傾向が分かれるが，短期間しかトラフィックが取得できない場合はユーザごとに差異を見つけるのは困難である．

- **接続位置**
  ネットワーク上の様々な位置に本システムを設置することによって，ホストの接続位置を取得する．これによって，ユーザの行動範囲を把握することが可能となる．しかし，この情報を利用できるかはネットワーク構成に依存するため，今回は取得しないものとする．

- **TCP SYN パケットの送信先 IP アドレス**
  ホストの送信先 IP アドレスの頻度はホストを識別することができるのか調査を行った．今回は，接続頻度の高い送信先 IP アドレス上位 5 位を取得した．また，同じネットワークの内での通信は除外する．調査したネットワークは筆者が所属する研究室でのネットワークであり，取得期間は，2009 年 6 月 25 日 19 時 45 分から 22 時 20 分までである．その結果を表 4.3 に示す．
  各期間における調査対象のユーザの上位 5 位のリストうち，ホスト A と同じ IP アドレス 1 つ，ホスト B に関しては 2 つ見られたが，ホスト C は同じ IP アドレスが見られなかった．これらの情報は，ネットワーク構成や，ユーザの利用頻度によって，この容易に変動してしまうため，上位のリストのみでホスト識別要素として利用することは困難である．

4.2 同一セグメント上のユーザと取得情報

ユーザは同一セグメントに接続すると，ブロードキャストやマルチキャストといったネットワーク全体に送信する情報を受け取ることができる．その中でも，ファイル共有をする際に利用する情報はホスト利用者を特定できる可能性がある．他にも，ホストの固有識別要素である MAC アドレスも取得できる．そこで，Windows や MacOSX で利用されている mDNS や NetBIOS と MAC アドレスなどホスト情報を利用することで，ネットワークにおけるユーザのプライバシを脅かす手法を提示した．第 4.1 節のパケットのヘッダ情報によるホストの識別と異なる点は，既にホストの識別された情報を利用するため，ホスト利用者のプライバシが脅かされる点である．

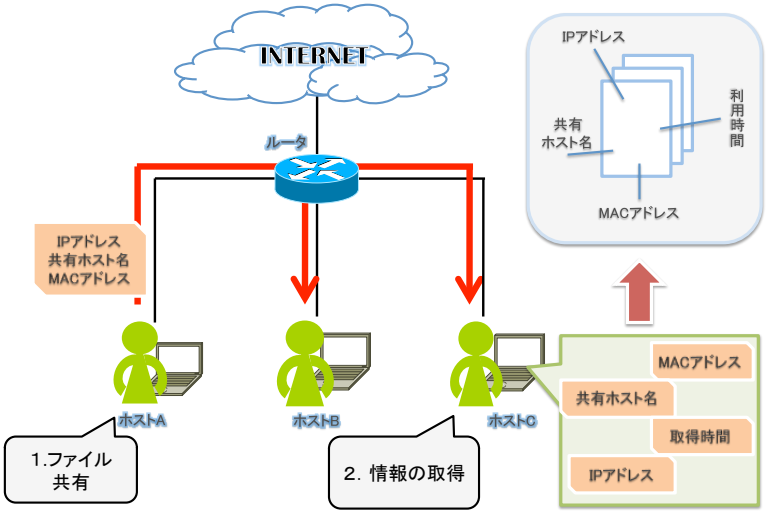


図 4.12: パケットヘッダ情報の収集

4.2.1 前提

前提として，スイッチなどで対象とするユーザの NetBIOS，mDNS といったプロトコルの取得を妨げることはないネットワーク構成を想定する．情報を取得するユーザは，一般ユーザであり，ネットワークにおけるサーバ，ルータ・スイッチ類へのアクセス権限はないものとする．また，無線 LAN であればプロミスキャス　モードを利用して，同じチャンネルに接続している情報を収集できるが，今回はその場合を想定せず，有線ネットワークを利用した環境を想定する．図 4.12 にその概要図を示す．

4.2.2 共有ホスト名

ネットワークに接続すると，OS によっては DHCP サーバやプリンタをはじめとする機器を探索する．探索はブロードキャストやマルチキャストといったネットワーク全体または複数のホストに送信される．その中でも，ファイル共有のプロトコルは共有ホスト名や OS 情報といったプライバシに関する情報が含まれている．これらの情報はネットワークのトラフィックに頻繁に見られる傾向にある．そのため，これらの情報を収集することで，ホストやユーザのプロファイルが作成できる可能性が高い．

mDSN や NetBIOS は設定した共有ホスト名や機器名をマルチキャストで送信している．そして，それらの情報は簡単に閲覧可能である．例えば，図 4.13 では，MacOSX の Finder アプリケーションのキャプチャーを表示している．ここで表示しているのはネットワークにおける共有時のホスト名を表示している．これによって，このネットワークにおいて，どのホストが共有可能かが判別できる．また，MacOSX の場合，デフォルトで共有が設定されている場合がある．その場合，ユーザ名とホストで表示される．図 4.13 にある，**ISC のホスト**等が挙げられているが，この場合ユーザが ISC という名前で設定されている．し

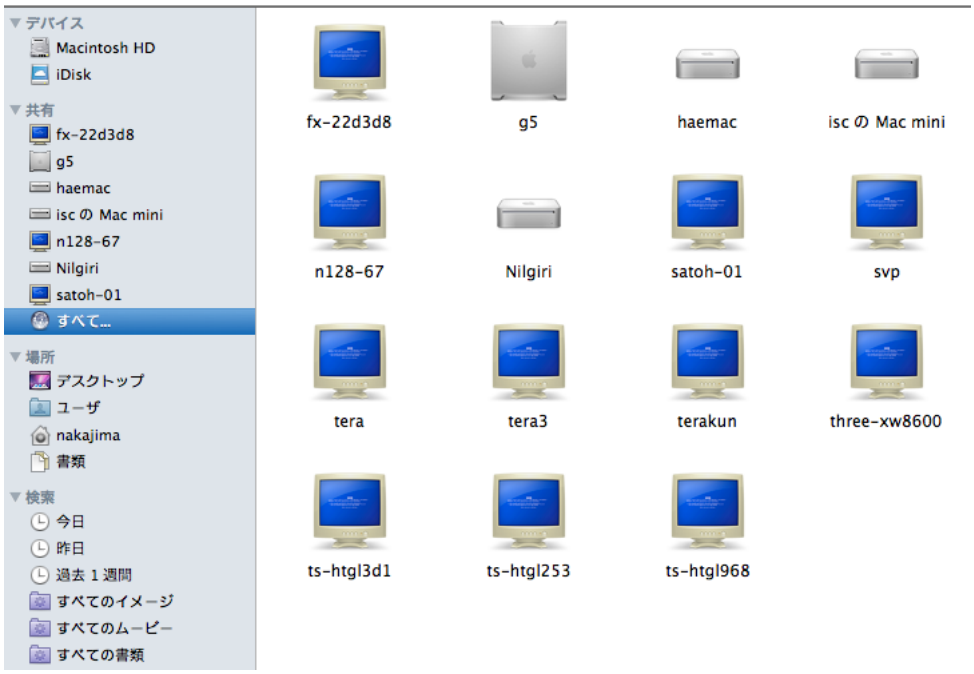


図 4.13: MacOSX の Finder

かし，ホストが個人所有のものである場合，**上原雄貴の Mac Book Pro** と表示されてしまい，ホストの所有者がネットワーク上で公開されている．

図にも示すように，同一セグメントに接続するだけで，共有を許可しているホストの機器名や名前を取得することができる．共有ホスト名だけでなく，機器固有の MAC アドレスも同時に取得できるため，それらを組み合わせてホストやユーザの特定が可能である．特に MAC アドレスといった一意情報と共有ホスト名を組み合わせると，ホストと人のマッピングすることができる．そして，この情報を蓄積し，利用することによって，ユーザの場所や生活時間など実生活上のプロファイル情報や場所情報，利用しているホストの OS といったプロファイルを作成することができる．また，一人で複数のホストを利用しているユーザであったり，グループで共有して保持しているホスト群を推測することができる．

これによって，ホストの利用者の実生活の情報やユーザが所持しているホスト数が分かる．そして，共有ホスト名と MAC アドレスを組み合わせることによって，ネットワーク上でユーザの追跡ができるようになる．

4.3 第三者であるユーザと取得情報

同じネットワークに存在しないユーザの取得できる情報は限定される．同じネットワークに存在しないユーザはトラッキング対象のユーザが利用するようなサービスを提供するか，ユーザが周囲にむけて発信する情報を取得のみできる．

ユーザが周囲に発信する情報に焦点を当てると，無線 LAN や Bluetooth の通信が挙げられる．その中で，第三者でも容易に取得可能な Bluetooth の通信に着目した．Bluetooth

はペアリング時に Bluetooth 名を周囲に発信するが，その情報がユーザ特定の識別要素になる可能性が高い．

4.3.1 前提

Bluetooth の情報を収集するにあたり，収集する機器は Bluetooth しかネットワークに接続することはできず，それ以外の情報を収集することはできないとする．また，収集する場所のばらつきを防ぐため，収集する機器は固定し，常に Bluetooth デバイスを探索し続ける．

4.3.2 Bluetooth

近年多くの機器に Bluetooth が導入されている．Bluetooth は簡単にユーザも発信することができ，周囲もそれを知ることができるため，Bluetooth デバイスアドレスが個人のプライバシに影響する場合は問題である．Bluetooth は機器によってはデフォルトで設定されており，ユーザは知らずに利用している可能性がある．また，ユーザが理解して利用している場合でも，プライバシが脅かされる可能性がある．そこで，Bluetooth によってホストだけでなく携帯電話など Bluetooth を利用する機器のプロファイルを作成する手法を提示する．

第**??**節で，Bluetooth のペアリングについて述べたように，ペアリングにおいて，機器の通信範囲全体に Bluetooth デバイスアドレスを送信する．そのため，ネットワーク上で機器固有の識別要素として扱われる可能性がある．実際に，携帯にデフォルトで設定されている場合があり，常に Bluetooth デバイスアドレスを発信し続けているケースが有る．そのような Bluetooth デバイスアドレスを取得して保持しておくことでユーザのトラッキングする手法を提示する．実際に，駅など人が集まるところで，Bluetooth デバイスを探索し，トラッキングをおこなったという試みもされている [21].

Bluetooth アドレスによって，機器を保持したユーザを推測できる．また，ホストのファイル共有とは違い，携帯電話など小型な機器までも Bluetooth に対応しているため，RFID タグのような利用も考えられる．これによって，実ユーザが移動した場所や滞在時間なども組み合わせてプロファイルを作成することができる．図 4.14 に示すように，各場所に本システムを設置することで，その人の訪問した時間，滞在時間，退出時間などを知ることができる．このシステムを要所に設置することで，実空間におけるのトラッキングが可能である．これによって，Bluetooth を利用しているすべてのユーザが影響を受ける．特に，実空間や移動した場所が記録として保持される．また，Bluetooth デバイスアドレスは固有で変更ができないため，一度プロファイリングされてしまうと，機器を変える以外に追跡を防ぐ手段がない．

以上のことからこの手法の検証は必要である．Bluetooth デバイスアドレスを収集し，ユーザの実生活や機器の ID を利用したプロファイル作成がどの程度の情報を収集できるのかを検証する．

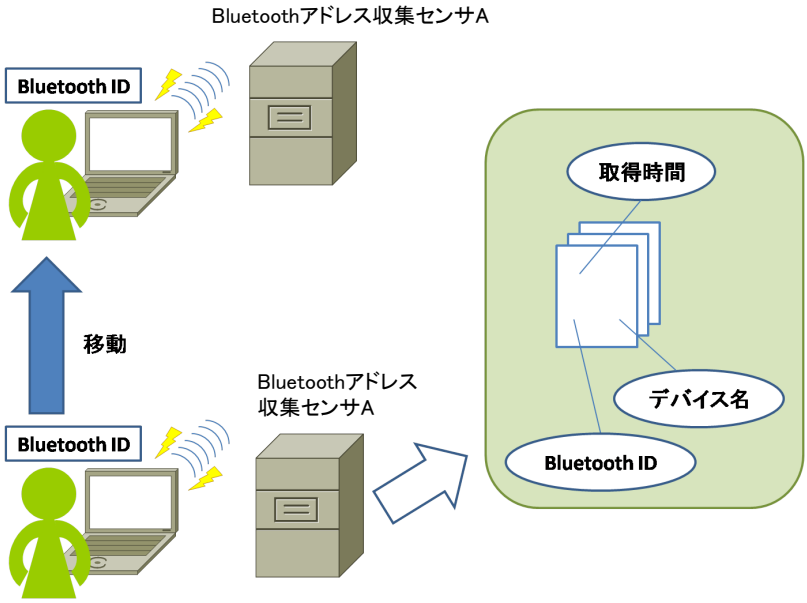


図 4.14: Bluetooth 情報の収集

4.4 検証情報統合によるリスク

検証に利用するパケットヘッダ情報，共有ホスト名，Bluetooth デバイス名を組み合わせることで，どのような情報を得ることができるのかを検討する．これらの情報を組み合わせることは個人を特定するにあたり脅威となる可能性がある．仮に，今回述べた情報を組み合わせることが可能になれば，ネットワークにおける個人をほぼ特定することができるようになる．もちろん，今回挙げた 3 つの情報だけではなく，それらの情報を取得するにあたり，利用した情報すべてを利用することでホストを利用するユーザを直接特定できる．

4.5 まとめ

本章では，ユーザのプライバシを脅かせる情報について 3 つの手法を提示した．まず，パケットのヘッダ情報だけをもとにユーザのプロファイルを作成する手法，ファイル共有時の mDNS プロトコルに利用される共有ホスト名，NetBIOS の NetBIOS 名を利用して，ホスト利用者の特定や実生活のリズムや位置情報を取得する手法，Bluetooth 探索におけるデバイスアドレスや名前を利用して追跡する手法の 3 つである．第 5 章ではこれらの手法を実際に検証する．

第5章　検証

本章では，第4章で述べたユーザ特定手法をそれぞれ検証する．また，検証に利用した情報を統合することによる脅威について述べる．

5.1　パケットのヘッダ情報

パケットのヘッダ情報を利用することでどこまでホストを識別可能なのかを検証する．取得する情報は主に5タプルと呼ばれる，送信先・受信元IPアドレス，送信先・受信元ポート番号，プロトコルである．

5.1.1　検証手法

ネットワークにおいてホストを正確に識別可能かの検証を行う．事前にターゲットユーザに同意を得てMACアドレスを取得する．ホスト識別にあたり，MACアドレスを使用しない．しかし，検証時にホスト識別の正誤判定に利用する．また，ターゲットホストのプロファイル情報をホスト利用者に提記することで，検証する．

検証手法の評価は，ホストを分類・識別の精度とする．そして，どのような条件下でホスト識別が可能なのか，もしくは取得に失敗した原因を究明する．ここでは，ホストに対するプロファイル作成を行い，過去のプロファイルした情報と合致した場合，同じホストであると判断する．

5.1.2　ホスト識別システム

ネットワークのトラフィックから各ユーザの送受信するパケット情報と接続位置や利用時間など情報を分析することで，ホストを識別する．ホストに関する情報は，ネットワークの中継地点に本研究を用いたトラフィック監視装置を設置することで，定常的に収集する．そのため，本システムの利用者は対象ネットワークの通信を監視する権限保有者，もしくはネットワーク管理者から許可されたユーザである．

システムの動作概要を図5.1に示す．ユーザのAliceは，本システムにおいてホストAとして推定されて，扱われる．そこで，ホストAから推測できる情報を利用してプロファイルを作成する．このプロファイル情報をもとに，ホストはトラフィックから位置情報やホスト情報などの情報を取得し，データベースに格納する．しかし，ホスト識別に利用する情報は，ホストの特徴やふるまいから作成されるため，Bobのような複数にまたがるホ

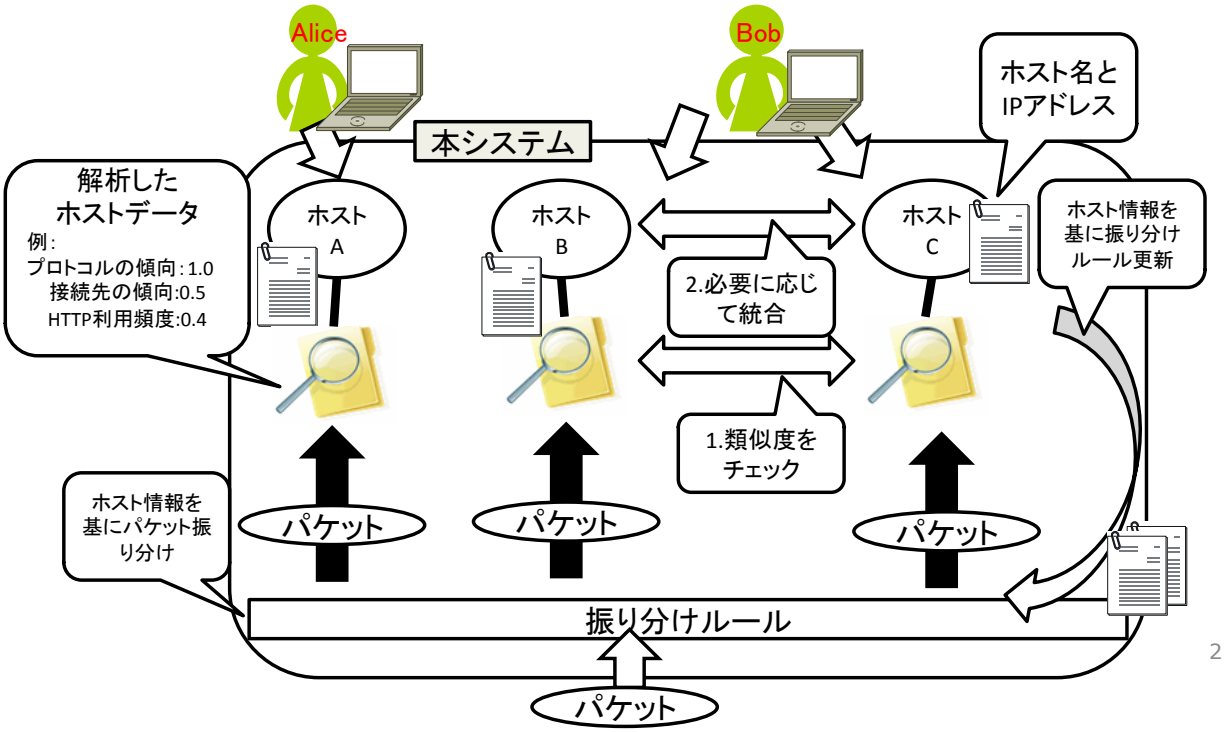


図 5.1: システムの動作概要

ストも想定する．その場合は，定期的にホスト間同士で共通事項を探索し，発見した場合は該当するホスト情報を統合する．

5.1.3 設計概要

本システムはネットワークの中継地点にトラフィック監視装置を設置し，定常的にパケットのヘッダ情報を収集することで，ホストを識別する．ホストを識別する本システムは，大きく 2 つの動作に分けられる．IP アドレスからホストを判定する **IP 判定モジュール**，ホスト上のアプリケーションの動作からホストを分類する**解析モジュール**である．ホスト識別手法の設計を図 5.2 に記す．

1. **IP アドレス判定モジュール**
   IP アドレス判定モジュールは，パケットの IP アドレスを基に，ホストを推定する．ネットワークの DHCP リースタイム時間内に該当する IP アドレスはすべて同じホストであると判断する．しかし，DHCP リースタイム時間内に該当する IP アドレスから通信がなかった場合は，そのホストはネットワークから離れたものとする．

2. **パターン解析モジュール**
   ホスト上のアプリケーションの動作からホストを分類する．このモジュールは状況に応じて複数個作成する．また各モジュールごとにパターンを解析するアルゴリ

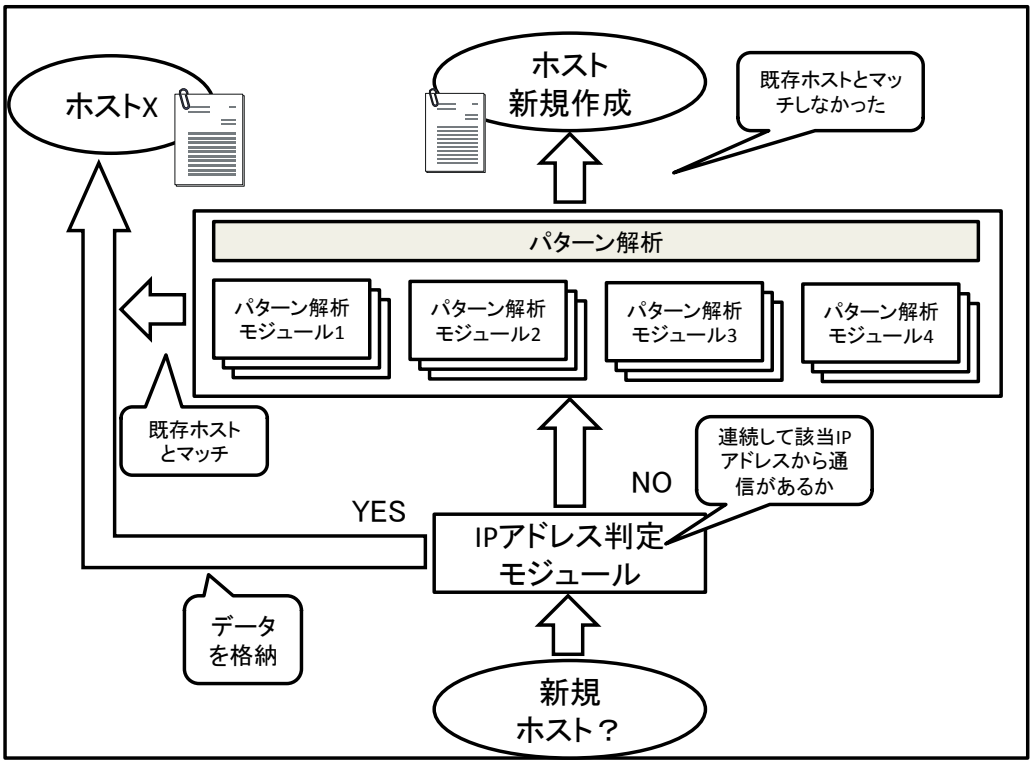


図 5.2: ホスト識別手法の設計

ズムは変化する．例えば，メールサーバにアクセスする頻度や回数をユーザを分類するための識別要素とする．

各モジュールで類似するホストを発見した場合，ホストが保有する情報は該当するホストの情報に統合される．そして，これらのモジュールによって新規ホストが既存のホスト情報とマッチしなかった場合にはじめて，新たなホストが作成される．これら一覧の作業を繰り返すことによって，ホストを識別する．

本システムの要件に対する充足度について述べる．本システムは，ネットワークの中継地点に設置するだけでホスト識別が可能になるため，管理者による運用が容易である．次に，ホスト識別の精度に関しては，本人と推測できる情報を多数組み合わせることによって精度の向上を図る．そして，定常的に情報を収集し，リアルタイムにホストを識別をすることができる．また，本システムは大規模ネットワークの上流で利用するため，ネットワーク上すべてのホストを対象とすることから，ユーザの網羅性はあると言える．また，モジュールの追加が容易である．

5.1.4 ホスト識別に用いる情報

ホストの識別に用いる情報は，ホストを推定できる情報である．推定に利用する情報は第 4.1.3 節で述べた以下の情報を利用する．

- **送信先 IP アドレスとポート番号の組み合わせ**
  送信先 IP アドレスとポート番号の組み合わせを調査する．利用するプロトコルは SSH，IMAP,VPN,IRC に限定する．このプロトコルで 3 回以上同じホストに対して通信が行われた場合，識別要素としてリストに加える．

- **発信元ポート番号**
  ホストの発信元ポート番号から OS を特定する識別とする．前述したように利用している送信元ポート番号から OS の種類をおおまかではあるが分類することができる．IP アドレスが付与されて 1 分間，対象ホストの発信元ポート番号を記録する．次に，発信元ポートが 1024-5000 か，49152-65535 どちらの範囲に近いかを判定し，識別要素の一部とする．また，ホストが休止状態から復旧した場合，連続したポート番号を利用するため，ホスト毎に最後に利用したポート番号を記憶する．

- **パケットの発信タイミング**
  パケットの発信タイミングによって，OS 情報を得ることができる．ホストを起動して IP アドレスが付与されてから 2 分間取得し，マルチキャストを除いた TCP の通信が最初に発生するタイミングを計測する．Windows の場合, IP アドレスが付与されてから 30 秒内に TCP による Update 確認の通信が発生することが確認されているため，判別が可能である．また，IP アドレスが付与されてから 5 秒以内に数十の TCP パケットが発信された場合，そのホストは，休止状態からの復旧，もしくはの電源をいれたままネットワークを移動したと判断する．

- **利用アプリケーション，サービス**
  サービスを提供するサーバのアドレスブロックを用意し，ホストが通信したサーバと照らし合わせることでホストが利用しているアプリケーションやサービスを特定する．また，ソフトウェアだけでなく OS も Update の通信も対象とする．

5.1.5 検証環境

2009 年 9 月 7 日（月）から 10 日（木）にかけて，長野県長野市松代町にて開催された WIDE 合宿におけるネットワークで実験を行った．合宿ネットワークはは 200 台以上のホストが参加しており，所属するコミュニティも様々である．その WIDE 合宿ネットワークトポロジ図を図 5.3 に示す．このネットワークですべての無線を利用したホストのパケットヘッダ情報を利用してホストの識別を行う．合宿でパケットを取得した期間は 2009 年 9 月 7 日 17 時 44 分から 9 月 9 日 16 時 8 分までである．

5.1.6 検証結果

ホストの分類

WIDE 合宿におけるネットワークから 6 台のホストを分類できるかという検証において，識別要素による手法を用いた．表 5.1 に，実験結果を示す．

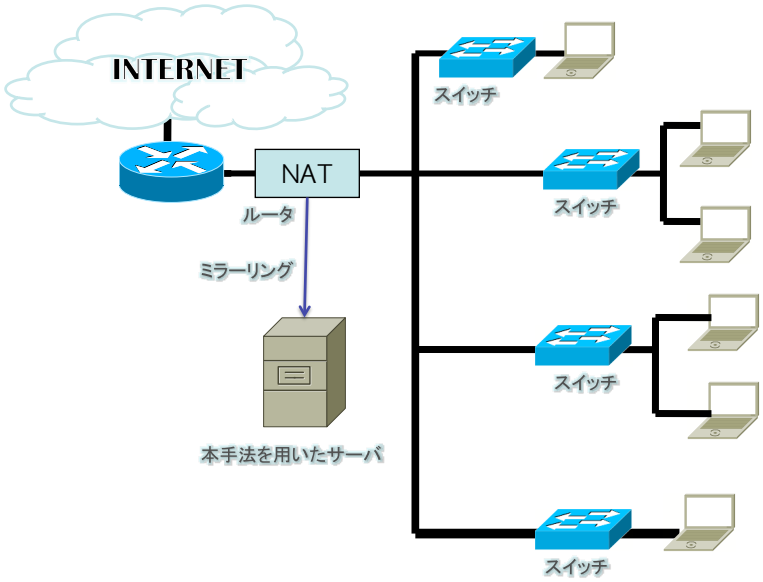


図 5.3: WIDE 合宿ネットワークトポロジ図

表 5.1: WIDE 合宿ネットワークにおける実験結果

識別要素＼ホスト	H_1	H_2	H_3	H_4	H_5	H_6
Mail Server	S_1,S_2 S_3	S_1,S_2 S_4,S_5	S_6	S_1,S_2	S_2	S_1,S_2
SSH Server	S_1,S_3 S_4,S_5	S_1,S_6 S_7,S_8 S_9	S_1			
パケット間隔から推測される OS	MacOS	MacOS	Windows	MacOS	Windows	MacOS
送信元ポート	49152-65535	49152-65535	1024-5000	49152-65535	49152-65535	49152-65535
IRC Server			S_1,S_2 S_3,S_4			
特徴的な挙動	定期的な HTTP 通信				特殊な 発信元 ポート利用	VPN, 定期的に HTTP 通信
MSN のリスト量	5383	1972	8862	13166		1362
起動時の挙動	HTTP 通信 IMAP 接続		HTTP 通信	MSN メッセンジャ	HTTP 通信	
Update	Evernote MacOS	MacOS	Windows	MacOS	Firefox Quick Time	MacOS Safari

H:ホスト **S**:サーバ

まず，SSH と IMAP における宛先ホストを分類した．特定のコミュニティに所属しているユーザは同じサーバを利用する傾向にあるため，宛先ホストを閲覧するだけで容易に分類が可能である．これによって，200 台以上のホストから 13 台まで絞ることができる．更

に，SSH，IMAP 利用者においてトラフィック受信量を識別要素とした場合，12 台のユーザの特徴付けを行った．SSH と IMAP の接続先で 12 台のユーザは各々に分類出来る．

IMAP の接続先に焦点を当てると，大きく 3 台の Mail サーバが利用されていることが分かる．更に詳しく解析すると，H_1,H_2は 多くのメールサーバとやり取りする傾向があるのに対して，H_3 は複数のアカウントを外部メールサーバで統合している可能性が高い．そして，SSH の接続先を見ると，H_1,H_2 は複数のホストを利用している．SSH と IMAP を利用するだけで各ユーザの分類は可能である．

次に，メッセンジャや IRC を利用しているユーザを分類した．メッセンジャを利用する場合，通信を行う相手のリストをやり取りする．そのリストのトラフィック量 (byte) をユーザごとに分けた．メッセンジャのユーザリストは増減するため，100byte 程度の幅を持っている．これによって，MSN メッセンジャを利用しているユーザすべての特徴を抽出することができた．また，ネットワーク会場の関係上，ホストをサスペンドをするユーザが多くいたため，復旧時の時間と送信元ポートを利用することによって，継続してユーザをトラッキングすることができた．

起動時のパケットとソースポート番号を観測することによって 5 台が MacOS であり，2 台が Windows であることが分かった．そして，ネットワーク接続時に OS やソフトウェアの Update を確認しているか監視した．その際に，FireFox を利用しているユーザが 4 名発見されている．また，アプリケーション利用時に発生するトラフィックから，Evernote[22] などを起動時に設定しているホストも判明している．

特徴的な通信について述べる．H_1 は起動時から一定時間ごとに同じサーバに対して HTTP でデータを送信する傾向がある．このことから，何らかのアプリケーションが起動していることがわかる．今回は，アプリケーション利用時の通信サーバのアドレスブロックを保持していると仮定しているため，このアプリケーションが Evernote であると判明している．同じく H_6 も起動時から同じサーバにアクセスし続けているため，常駐アプリを自動起動に設定していると言える．H_5 では発信元ポート番号の傾向が 49125-65535 番を利用する OS であるが，発信元ポート番号 1444 を何度も連続して常に利用している．このため，特殊なサービスを利用していると推測できる．また，起動時に毎回 HTTP を利用したダウンロードしているのが観測された．H_5 は起動時のパケットのタイミングから Windows であることが判明しているため，アンチウイルスソフトのシグネチャの更新である可能性が高い．実際に通信先 IP アドレスから Avast と呼ばれるアンチウイルスソフトを利用していると判断された．H_6 では，VPN を利用するとともに起動時から定期的な HTTP 通信をしている．これら識別要素をもとにホストを分類することで，ホストごとの特徴を分けて表すことができる．それらの識別要素を，分類した結果 6 台中 6 台すべてを一意に分類することができた．

ホストの識別

ホストを分類することで得た識別要素を利用して，ホスト識別する．識別に必要な情報は Mail サーバもしくは SSH サーバへのアクセスである．これによって，200 台以上あるホスト群から 10 台前後に特定出来るため，この接続先とポート番号のセットは不可欠

である．その識別要素と更に別の情報を加えることで初めてホストを個別に識別できる．ポート番号と接続先の組み合わせ以外には，MSN メッセンジャのトラフィック量が有用である．これは，メッセンジャの場合，連絡先のリストの数はユーザごとに異なる可能性が高いため，ホスト識別に利用しやすい．また，起動時の挙動から，一定のホストにアクセスをするために，利用アプリケーションを要素とすることで識別できる．

5.1.7 考察

宛先ホスト，プロトコル，利用アプリケーション，起動時の挙動，パケットのタイミングをもとに，すべてのユーザを識別することができた．IMAP や MSNMS のトラフィック量からホストを識別する手法は有効である．また，サスペンドからの復旧を追跡するために送信元ポートを利用する手法も有効であることが示せた．そして，継続的にホスト追跡することで，アプリケーションのアップデートや固有の通信を得ることで，ホストのプロファイル作成ができる．このように様々な情報を組み合わせることで，パケットのヘッダ情報だけでもホストの識別は可能である．

しかし，IMAP や MSN メッセンジャのプロトコルは，ユーザの利用頻度や HTTP を通じて利用されるなど，ユーザの振る舞いに大きく変化していしまうため，識別要素としてなり難い．また，サービスを特定するために，サービスするアドレスブロックを保持しているが，アドレス範囲が不明な場合，本システムの識別要素として利用している情報が取れない場合があることが分かった．今後，精度をあげるために，より多くの識別要素を追加する必要がある．

5.2 共有ホスト名

同じネットワークにおけるユーザが共有ホスト名を取得するだけでユーザのホストに関する情報が取得できるかを検証する．取得する情報は SMB と mDNS プロトコルである，共有ホスト名と MAC アドレスである．それらの情報を用いて，ユーザの実生活やホストの OS 情報，ホストの所有者を推測する．

5.2.1 検証手法

検証手法としては筆者が所属する研究室のネットワークに本検証手法を用いたホストを接続し，情報を取得し続ける．取得した期間は 2009 年 11 月 18 日から 20 日の約 3 日間である．取得したホスト名をもとに，ユーザのホスト情報や実生活での情報を結びつける．そして，ホスト名と MAC アドレスを結びつけて，リスト化して保存する．

表 5.2: 共有ホスト名の手法検証の実装環境

仕様言語	version	ライブラリ	OS
C 言語	4.2.1	libpcap	FreeBSD7.2
PHP	5.3		MacOSX 10.6.2

5.2.2 設計概要

検証を行うプログラムは表 5.2 で挙げるように libpcap によるパケットキャプチャを C 言語で記述する．プロトコル名とポート番号で NetBIOS と mDNS を判別し，ペイロードからホスト名のみを抽出する．NetBIOS は自身の共有ホスト名しか発信しないが，mDNS の場合はホストが持っている共有ホスト名のリストを送信するため，まとめて取得する．SMB の場合は共有ホスト名と MAC アドレスの結びつけは容易であるが，mDNS によるリストから共有ファイル名を取得した場合は，MAC アドレスとの結びつけが難しい．そこで，リストから共有ファイル名を取得した場合は，リストに記載された共有ファイル名のホストが通信するまでトラフィックを監視し，MAC アドレスが取得できるようになれば，情報を組み合わせる手法を採用する．そして，一度組み合わせた情報は保存する．

トラフィックから情報を取得後，FreeBSD 上で処理し，出力結果を PHP で記述したスクリプトによって，グラフ化した．

5.2.3 検証環境

取得範囲はスイッチによる情報制限を受けないため，研究室全般のネットワークから情報を収集可能である．そこで，研究室のネットワークにおいて，2009 年 10 月 18 日から 21 日まで本システムの実験を行った．図 5.4 に共有ホスト名を用いた実験のネットワーク概要図を示す．

5.2.4 検証結果

本システムによって，共有ホスト名とホストの MAC アドレスを結びつけた．検証の結果，取得した共有ホスト名の数は 28 台であった．常に稼働しているホストは 8 台あり，20 台のうち同じユーザと推測できるホストが 7 台あった．

次に，共有ホスト名をもとに，ユーザの生活時間帯を取得することができた．図 5.5 にその結果を示す．このグラフは計測時間を 30 分ごとに区分けして，ある時間でホスト A が発見された場合，ホスト A は観測された時間帯はネットワークにいると仮定する．プロトコルの関係上，一定時間ごとにリストの情報をやりとりをしないため，この手法を採用した．

この結果から，大まかなユーザの生活リズムの特徴が導き出される．ユーザ A は常に研究室に在籍しており，高頻度で検出されている．次に，ユーザ B は夜のみ研究室で観

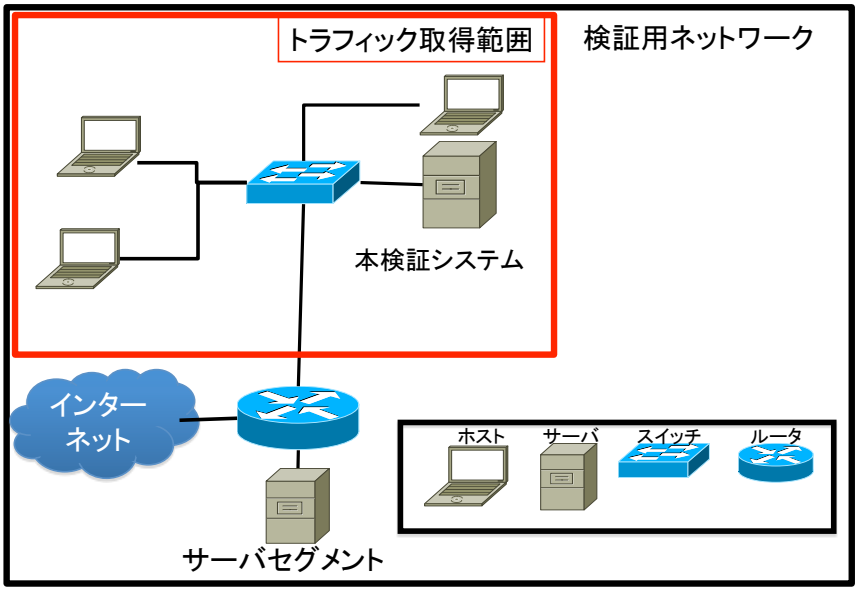


図 5.4: 共有ホスト名を用いた実験のネットワーク概要図

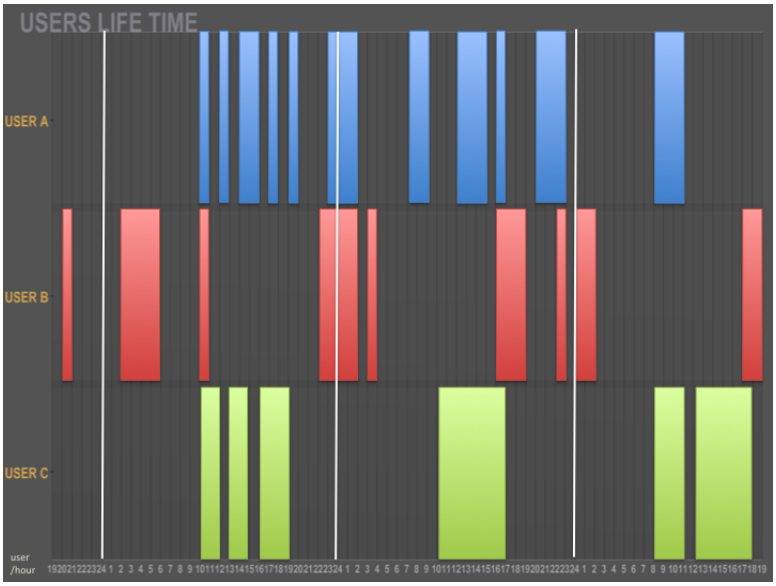


図 5.5: 共有ホスト名を用いたユーザ生活モデル

測されているため，夜型の生活を送っていることが推測できる．ユーザ B とは逆にユーザ C は朝方の生活を送っており，この期間内は規則正しく研究室を出入していることが分かる．

表 5.3: Bluetooth デバイスアドレス検証手法の実装環境

仕様言語	library	OS
java1.6	Bluecove 2.1.0	MacOSX 10.6.2
PHP		MacOSX 10.6.2

5.2.5 考察

共有ホストのシステムでは，ユーザとホストを結びつけることが容易になることを示した．本システムによって，MAC アドレスと共有ホスト名の結びつけをすることで，複数ホストを利用している場合や，グループで管理しているホストの情報を収集できる．このシステムによってネットワーク上でファイル共有を利用しているホストの特定が可能であるため，ユーザのプライバシを脅かす可能性が非常に高いと言える．この情報を保持しておくと，MAC アドレスとホストの対応が分かるため，サーバのログや，パケットのヘッダ情報と組み合わせることで，より詳細なプロファイルが作成できる．ファイル共有プロトコルは iTunes[23] をはじめとする多くのアプリケーションに利用されているため，取得が容易であると言える．

ただし，NetBIOS や mDNS といったプロトコルはサービスを利用していない場合でも送信し続けている場合があるため，今後も仕様や設計も調査する必要がある．しかし，ファイル共有を利用していないホストの識別が不可能である．また，共有プロトコルは常に送信し続けるわけではない．そのため，分単位でのユーザの生活時間を取得できないという欠点がある．

5.3 Bluetooth デバイス名

5.3.1 検証手法

筆者が所属する研究室において，Bluetooth を検出するプログラムを実行することで，どの程度情報収集できるかを検証した．実際に，ペアリング時に Bluetooth アドレスを収集し，ユーザの実生活や機器を利用したプロファイル作成がどの程度の情報を収集できるのかを検証した．

5.3.2 設計概要

Bluetooth アドレスの検証実験を行うプログラムは表 5.3 で示すように，java 言語で記述し，MacOSX 上で実行する．使用するライブラリは bluecove 2.1.0 を利用し，Bluetooth デバイスの探索を行う．デバイス探索で検出される情報は，取得した日時，Bluetooth アドレス，Bluetooth デバイス名である．デバイス探索の間隔は 30 秒に一回の割合で連続し

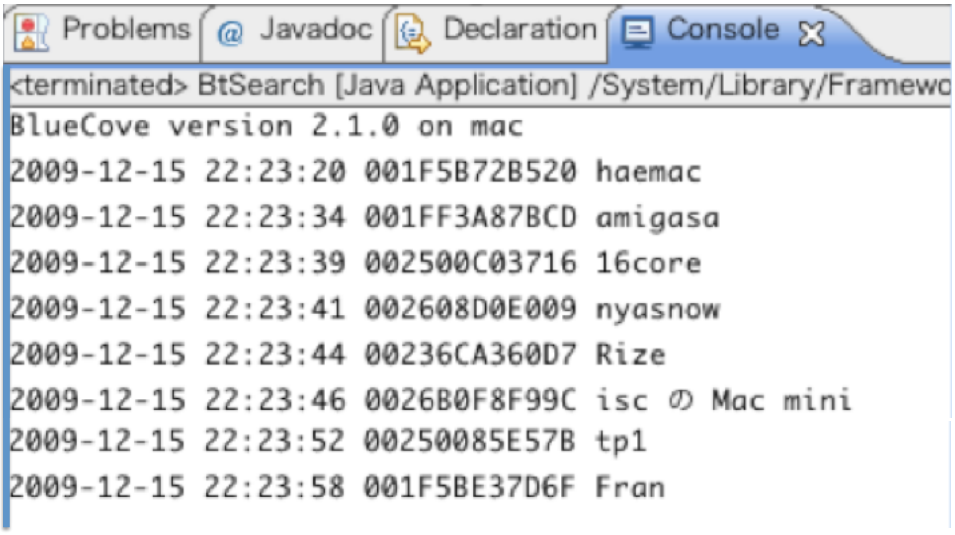


図 5.6: Bluetooth デバイスの検出

表 5.4: Bluetooth デバイスアドレス検証結果

発見ホスト数	常駐ホスト数	携帯端末数
26	5	3

て行う．デバイス探索後，取得した情報をファイルに出力し，出力したファイルを PHP によるスクリプトによって，グラフによる視覚化をする．

5.3.3 検証環境

検証環境は筆者が所属する研究室だが，取得する範囲は筆者の実験ホストから周囲数メートル範囲である．データを取得した期間は，2009 年 12 月 17 日 23:00 から 2009 年 12 月 19 日 21:00 までである．Bluetooth デバイスを探索するホストの設置場所は研究室の中心部である．Bluetooth の有効範囲は数 m から数百 m と機器や環境によって差があるため，場所やデバイスによっては検出できない場合がある．今回は MacBook Pro と ThinkPad T41 の 2 台を用いて，事前に検証を行った．研究室は縦 22m，横 7m の広さであるが，どの場所でも Bluetooth デバイスが検出されたため，手法検証を行う上で本実験機器の配置場所を問題ないとする．

5.3.4 検証結果

上記のプログラムを実行したところ図 5.6 のような情報を収集した．ここで取得できる情報は Bluetooth デバイスアドレスと取得時間，設定している Bluetooth デバイス名であ

る．共有名と同じく，MacOSX を利用しており，かつデフォルトで設定をしていない場合は所有者とホストの機器名が表示される結果になった．

次に，取得したデータを解析した結果を表 5.4 に示す．検証期間内に取得した Bluetooth アドレスは全部で 26 あった．第??節では，この研究室に 100 台近くのホストがあることを示しているため，Bluetooth を利用しているホストは比較的少ないといえる．その中で，24 時間常に稼働しているホスト 5 台発見された．これらはサーバとして利用されている可能性が高い．次に，ホストではないスマートフォンと思われる機器が 3 つ観測されている．また，共有ホスト名に表示される情報をもとに，ホスト間の関係が推測ができる．今回観測された情報から，Bluetooth デバイス名が同じ機器が 3 組あった．そのうちひとりで 3 台のホストを利用しているユーザや，グループで管理されているホスト郡が観測された．このように，Bluetooth デバイス名を利用するだけでユーザやグループがどのくらいのホストを利用しているのか推測できる．

また，Bluetooth アドレスは固有のものであり，ホスト名とバインドすることによって，共有ホスト名と同じように，生活時間や場所のトラッキングが可能となる．図 5.7 に取得した情報によって作成したグラフを示す．このグラフの作成も，図 5.5 と同じく手法で作成した．Bluetooth デバイス検索は常時行われるため，連続して収集することができるため，5 分毎に区分けし，観測された時間帯は収集機器の周辺にユーザがいるものとする．

この検証では，一般的に持ち歩きしているユーザのホストを 3 種類選び出した．3 人のユーザも特徴が出ている．ユーザ A，B は比較的夜型の傾向があることに対して，ユーザ C は昼に研究室に訪れている．また，この検証は木曜日から土曜日にかけて行っているため，どのユーザも土曜日の夕方まで研究室を訪れてないことが分かる．

5.3.5 考察

Bluetooth のシステムを利用した結果，Bluetooth デバイス名と Bluetooth アドレスを取得することができた．このシステムを応用することによって，ユーザの場所や生活時間を取得することができる．ユーザの場所は，本システムを様々な場所に設置することによって検出可能である．本システムでは 30 秒に 1 度デバイス探索をするため，共有ホスト名のシステムよりも容易かつ正確に取得できる．ターゲットが学生であり，設置場所が研究室だと想定すると，ユーザが検出されなかった時間の講義やイベントを探すことで，ユーザがどこにいるのかを推測することもできる．このことから，Bluetooth はユーザプライバシを脅かす可能性があるといえる．Bluetooth デバイス名は設定していない場合はユーザ名と OS 名となる．そのため，共有ホスト名のシステムと連動することで，MAC アドレス，ホスト名，Bluetooth アドレスの 3 つを結びつけることができる．そして，結びつけた MAC アドレスから共有ホストと同じように多くの識別要素と結びつけることで，ユーザのプロファイルが可能となる．検証の章でも述べたように，Bluetooth を利用するユーザが多いとは言えないが，今後の Bluetooth の普及次第によっては非常に強力な識別要素となる．しかし，Bluetooth のデバイス探索は，ユーザが探索をしたい場合にのみ利用するのが一般的であるが，実験結果や高木浩光の調査で挙げているように常時探索をし

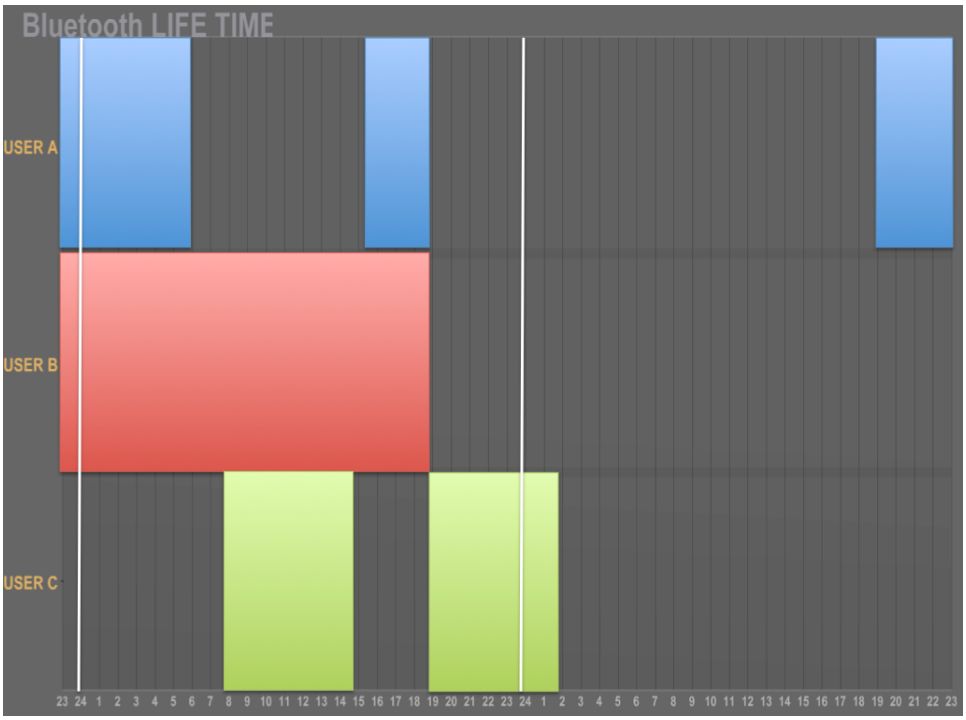


図 5.7: Bluetooth デバイスの検出によるユーザのライフタイム

表 5.5: 検証で利用した情報

利用情報	パケットヘッダ情報	共有ホスト名	Bluetooth デバイス名
一意にホストを識別できる情報	なし	MAC アドレス	Bluetooth デバイスアドレス
ホストの推測に利用できる情報	パケットのヘッダ情報 ホスト情報 利用サービス	共有ホスト名 位置情報 接続時間	Bluetooth デバイス名 位置情報 生活リズム

ているデバイスが数多くあることが分かる．常時探索が有効な機能であるかは議論が必要である．

5.4 検証した情報の統合

Bluetooth デバイス名，共有ホスト名，パケットのヘッダ情報を統合することによって，より正確な個人を特定することができる．検証に得られた情報を表 5.5 に示す．

検証の結果，個々の情報収集手法からホストの情報だけでなく，ホストや所有者の情報を取得できることが分かった．そこで，パケットのヘッダ情報，共有ホスト名，Bluetooth の情報を組み合わせるシステムを ISP やネットワーク管理者が利用した場合，ユーザが特定される可能性を考察する．本手法では，パケットのヘッダ情報とその傾向を利用して

プロファイルを作成する．そのため，ISP のような大規模ネットワークにおいて，パケットのヘッダ情報といった識別要素のみではユーザのプロファイルが困難である．識別要素には，共有ホスト名や，Bluetooth のデバイスアドレスも識別要素として利用可能である．パケットのヘッダ情報だけでは，ユーザの振る舞いによって，ホスト識別の精度が変化する．しかし，共有ホスト名と MAC アドレス，Bluetooth のデバイスアドレスを組み合わせた情報を保持することで，条件が整えばユーザがホストを複数持っていた場合や，ホストを変えても追跡が可能である．最終的には，ホストと携帯電話などの機器の組み合わせ，ネットワークに繋がるすべての機器と，Bluetooth を利用できる機器を結びつけてプロファイルすることができる．そして，各機器から，実ユーザの位置，時間，行動傾向といった個人に関わる情報が本人の知らない間に利用される危険性がある．

以上のことからパケットのヘッダ情報，共有ホスト名，Bluetooth デバイスは，ユーザをトラッキングするにあたり，非常に有効な手段であるとともに，ユーザのプライバシに脅威を与えるものであると言える．

5.5 まとめ

本章では 3 つの仮説を検証した．パケットのヘッダ情報と，共有ホスト名，Bluetooth によるホスト識別の可能性である．パケットのヘッダ情報のみを用いて，ホストを識別する手法は，ネットワークの中継地点に設置し，定常的にヘッダ情報のみを収集し，識別要素をもとにホストを分類する．これによって，ネットワーク上のホスト 7 台のうち 4 台を識別することができた．次に，共有ホスト名，Bluetooth アドレスはホストとユーザ名をデフォルトで設定しているとユーザが利用しているホストやユーザ自身の生活の推測が可能になることを示した．これら 3 つの情報は，ユーザのプライバシを脅かす可能性が十分にあると言える．

第6章　ガイドラインの提案

本章では第5章での検証結果をもとに，ガイドラインの提案した．ガイドラインは一般ユーザと開発や管理者に分けて提案する．パケットのヘッダ情報，共有ホスト名，Bluetoothデバイス名はユーザのプライバシを脅かす可能性があることは前述した通りである．しかし，サービスを受けるためには，これらの情報は利用しなければならない．そこで，これらの情報を扱う際のガイドラインを提案する．ガイドラインの対象は主に，一般ユーザ，開発者・管理者である．

6.1　一般ユーザのガイドライン

デジタルデバイスや情報の増加において，一般ユーザのプライバシ保護と利便性はトレードオフの側面がある．特に，ユーザのプライバシはユーザ自身も守る必要があるため，ユーザ自身のガイドラインが必要となる．一般ユーザの場合のガイドラインを図6.1に示す．

まず，ユーザを特定する識別要素となるデバイス探索機能はデフォルトでオフにし，常時探索をしない．サービスを受ける場合以外に利用しないよう設定する．これは，Bluetoothのデバイス名で挙げられるように，ユーザが意図せずに自身の情報を周囲に発信している場合があるためである．

次に，固有の識別要素を利用するサービスでは，利用時以外は使用しないことが重要である．サービス利用時以外にその機器を利用しないことは，周囲に自身の情報を発信する危険性を減らすことにつながるためである．

そして，ユーザはサービス利用時において，どのような情報が利用されているかを確認し，ユーザ自身が利用している機器のサービスを把握する必要がある．自身が発信している情報を知ることで，ユーザのプライバシに関わるかどうか判断することができる．また，サービスを受けるにあたり，ユーザの情報を受け取る先が信頼できるかどうかを確認しなければならない．例えば，個人情報の取り扱いの基準を満たしているかなどである．

これらのガイドラインを守ることによってユーザは自身のプラバシを保護することができると言える．ユーザ自身が発信している情報を知り，発信する情報をユーザが決めることで，ユーザのプライバシが脅かされる可能性を低減することができる．

6.2　開発者，管理者のガイドライン

開発者やネットワーク管理者のガイドラインを図6.2に述べる．

- デフォルト設定の確認
  利用端末が本人設定していないのに関わらず，識別要素を発信している場合があるので，設定を確認する．
- 目的とした場合以外ではサービスを利用しない
  サービスを利用していない場合は，サービスに関わる機能を使わない．
- 自身の利用するサービスの信頼度の確認
  サービスを受ける場合，個人情報を送信する相手が，どのように情報を管理をしているか確認する．
- 利用情報の確認
  サービスを受けるために，発信する情報が問題がないかを確認する．

図 6.1: 一般ユーザのガイドライン

- 情報統合におけるユーザの同意
  パケット情報とホスト名といった複数の個人情報を用いる場合は，必ずユーザの同意を得る．得ることが困難な場合は事前に概要を説明し，オプトアウト形式を採用．
- ユーザによる利用情報の選択
  どこまでの情報を利用して良いのかをユーザに選択を求める．
- 利用識別要素の限定
  情報を組み合わせる場合，Mac アドレスなどホストを一意に特定でき，変更が困難である識別要素を利用しない．
- 情報の公開
  個人情報を取得する際は，何の情報を利用するのか，何と組み合わせるのか，その結果，どういう情報が取得できるのかを明確にする
- ユーザによる情報管理
  ユーザに識別要素を付与もしくは利用する場合は，ユーザ側で容易に識別子の変更・削除といった管理できるようにする．
- 漏洩時の対策の明記
  ユーザの識別要素を流出してしまった場合の対策を明記する必要性．

図 6.2: 開発者，ネットワーク管理者のガイドライン

第 2 章で述べたように，ユーザに関係する情報を利用する場合はユーザの同意が必要である．ユーザの同意なしに情報の収集や改変を行った場合，プライバシを侵害する可能性がある．今回取り上げた 3 つの情報はプライバシを脅かすため，情報を収集・利用する場合は事前に同意を得る，もしくはオプトアウトの形式を採ることが望ましい．しかし，NebuAd での事例のように，法律に触れていないのにも関わらず，ユーザやプロバイダが情報収集されることを敬遠することがある．そのため，ユーザ，サービス提供者双方の利益のために，これらの情報を利用する際は，ユーザの同意を得ることが必要である．また，これらの情報を MAC アドレスのような一意性の強い情報と結びつけた場合，個人を特定されやすいため，プライバシ保護の観点から結びつけてはならない．MAC アドレスを利用することで，多くの情報を結びつけることができるのは第 6 章の冒頭で述べた通りである．

ユーザが識別要素を取得管理するモデルの例でも挙げたように，識別要素やそれに準ずるものをユーザに付与する場合，ユーザが自身で情報を発信している場合では，同意のほかにも，容易に管理できることが必須である．また，ユーザ自身が発している情報を識別要素として使用する場合は，その情報をユーザ自身が管理できるように告知することを提案する．

本ガイドラインで最も重要な点は，情報を収集を行う場合は利用目的，結果を明文化しユーザの同意を得ることと，識別要素となる情報は，ユーザ自身が自由に管理できるように情報を提供することの 2 点である．

6.3 ガイドラインの充足度の検討

前述したガイドラインにおける充足度を検討する．まず，ユーザのガイドラインについて述べる．ユーザの利用する機器はプライバシに配慮されている必要があるが，サービスを受けるためには，自身の情報を発信しなければいけない場合があり，利便性とプライバシのトレードオフである．今後はユーザ自身がサービスと利用される情報を理解し，同意をするという行為に重きがおかれることが予想される．そのためにも，ユーザの意識向上は不可欠であるため，本ガイドラインでは強調している．

管理者，開発者のガイドラインの充足度は OECD8 原則 [24] をもとに充足度を検討する．OECD8 原則は，1980 年に OECD に採択された個人情報保護に関する国際的なガイドラインである．OECD8 原則を図 6.3 に示す．OECD8 原則と本論文の提示するガイドラインを比較し，対応した結果を表 6.1 に記す．

まず，ユーザの同意に関する項目は，OECD8 原則の 1，情報制限の原則，データ内容の原則，目的明確化の原則，利用制限の原則と一致する．ユーザの同意を得ることは，情報収集においてはなくてはならない前提である．次に，ユーザの利用情報の選択は，ユーザの同意が前提であり，同意しない場合は収集しないという条件に一致する．基本的にサービスはユーザのオプトイン形式をとることが望ましい．そして，識別要素の限定に関しては，OECD8 原則に該当する記述はない．しかし，ユーザのプライバシを保護に配慮するためにも必要であると言える．情報の公開に関しては，OECD のデータ内容の原則，

1 収集制限の原則
　個人データは，適法・公正な手段により，かつ情報主体に通知または同意を得て収集されるべきである．

2 データ内容の原則
　収集するデータは，利用目的に沿ったもので，かつ，正確・完全・最新であるべきである．

3 目的明確化の原則
　収集目的を明確にし，データ利用は収集目的に合致するべきである．

4 利用制限の原則
　データ主体の同意がある場合や法律の規定による場合を除いて，収集したデータを目的以外に利用してはならない．

5 安全保護の原則
　合理的安全保護措置により，紛失・破壊・使用・修正・開示等から保護すべきである．

6 公開の原則
　データ収集の実施方針等を公開し，データの存在，利用目的，管理者等を明記するべきである．

7 個人参加の原則
　データ主体に対して，自己に関するデータの所在及び内容を確認させ，または異議申立を保証するべきである．

8 責任の原則
　データの管理者は諸原則実施の責任を有する．

図 6.3: OECD8 原則

目的明確化の原則，利用制限の原則，責任の原則に一致する．何の情報を利用して，どのような情報が分かるのかを明確化することはユーザの同意ともつながるため必要不可欠である．ユーザによる情報管理は個人参加の原則と一致する．特に異議申し立てより，ユーザ自身がいついかなる場合でも削除，更新できるように管理できるモデルが望ましい．漏洩時の対策の明記は，安全保護の原則，公開の原則と一致する．どのように守るのかだけではなく，漏洩時の対策について明記するべきである．

このように，プライバシのガイドラインとされた OECD8 原則において，すべての条件を満たすだけではなく，新たな要素を盛り込んでいるため充足度を満たしていると言える．

表 6.1: ガイドラインの充足度

ガイドライン	対応する OECD8 原則	備考
ユーザの同意	1,2,3,4	
ユーザに利用情報選択	1	
識別要素の限定		該当欄なし
情報の公開	2,3,4,8	
ユーザによる情報管理	7	
漏洩時の対策の明記	5,6	

6.4 まとめ

一般ユーザのプライバシを保護するために，個人情報を取り扱いに関するガイドラインをユーザとネットワーク管理者，開発者に対象を分けて提案した．一般ユーザ向けのガイドラインは，ユーザ自身が発信する情報について理解し，管理する必要性についての項目を設けた．次に，管理者，開発者向けのガイドラインには，ユーザの同意を得ることと，ユーザが自由に管理することに重点を置いた項目を設けた．そして，本ガイドラインの充足度について項目ごとに検討，考察した．

第7章　結論

本章では，本論文の成果をまとめ，第 1.2 節で示した目的の中で，達成された部分を述べる．そして，本論文における目的を実現するために今後の展望を述べる．

7.1　まとめ

本論文の目的はユーザが自身に関わる情報を管理することで，ユーザのプライバシが容易に脅かされない社会を実現することである．そのために，どのような情報がユーザのプライバシに関わる情報であるか議論が必要である．

情報収集技術の発展によって，プライバシの脅威が増加している．そのため，ISP をはじめとするネットワーク管理者は，どのような情報がユーザのプライバシを脅かすのかを改めて議論しなければならない．同時に，ユーザもどのような情報を発信しているか知ることで，自身のプライバシを保護する必要がある．

そこで，本論文では，デジタル情報収集において，ユーザのプライバシを脅かす可能性がある情報を用いた手法を 3 つを提示し，検証を行った．個人に関わる情報は多くあるが，情報収集者はユーザが発信する情報をすべて取得できるわけではない．ユーザと収集者のネットワーク上の関係によって異なる．そこで，ユーザが発信する情報を情報を収集する者が同じネットワークにいない場合，同一セグメントの場合，ネットワーク管理者の場合に分けて情報を分類した．

各々の場合において，ユーザが定常的に発信している情報を元に，プロファイル作成の手法を提示した．利用する情報はパケットのヘッダ情報，サービス探索情報，Bluetooth デバイスの探索情報である．これらの情報を利用したシステムを作成し，検証を行った．パケットのヘッダ情報を利用する手法では，パケットのヘッダ情報のみでホストを識別できるかを検証した．検証の結果，6 人中すべてのホスト識別することができた．同時に，ユーザのアクセスする傾向にあるサイトや，利用しているアプリケーションと言ったプロファイルの作成ができた．次に，共有ホスト名を利用した手法では，共有ホストの名前と MAC アドレスを結びつけることによって，ネットワークにおいてユーザのプロファイルが可能になることを示した．また，複数のホストをもつユーザやグループで管理しているホストの特定もできるという結果になった．最後に Bluetooth を利用した手法では，Bluetooth のデバイス名とアドレスを取得し，ユーザの生活時間や，場所情報を容易に取得できることが分かった．これら 3 つの情報はユーザを追跡する識別要素になることを示した．そして，3 つの情報を組み合わせることで，ネットワークに繋がるすべての機器と，Bluetooth に対応している機器を把握できることを示した．これらは情報収集者にとって

は有効なユーザの識別要素であるが，ユーザのプライバシを保護するために措置が必要がある．そこで，情報を扱うガイドラインを提唱することで，ユーザのプライバシを保護する手法を提案した．

本論文によって，ユーザが知らずに発信している情報を利用することでユーザのプラバシが脅かされる可能性を示した．これによって，いままであまり注意を払わなかった情報がプライバシを脅かす可能性があることが判明したため，再度ユーザのプライバシのあり方を議論しなければならないことを示した．そして，個人情報を含む情報の取扱を記したガイドラインを提案することで，新しいプライバシのあり方の一つを示した．

7.2 今後の展望

本論文は，デジタル情報収集におけるプライバシを保護する対策の一部分を示したにすぎない．デジタル情報においてプライバシと密接に関わる情報は膨大である．そこで，本論文で提案したガイドラインを拡張し，より多くの事例に対応する必要がある．例えば，開発者・運用者は個人情報を利用するにあたり，何の情報を取得するか，どの情報と組み合わせるか，その結果何の情報が取れるのかを明文化することが挙げられる．しかし，識別要素の組み合わせによるプライバシ侵害の可能性は抽象関数的に増加する．そのため，識別要素の特徴を抽出し，カテゴリ別に分類することで，カテゴリごとの組み合わせにより得られる情報の調査や分類手法の検討をしなければならない．そして，ガイドラインを正しく評価し，十分に広めるために，本論文における 3 つの情報だけではなく，ガイドラインを利用するケースについても考慮する必要がある．ガイドラインを適応する場所が，大学間や一般ユーザと企業，企業間によって，ガイドラインも全く異なる．そのため，様々な状況を想定しなければならない．ユーザのプライバシを保護するためには，サービス・コンテンツ提供者の側とユーザ側の双方から，情報を取得する手法を検討することが求められる．ユーザのプライバシ保護とサービスを受けることはトレードオフであるため，今後，情報を収集する側とされる側の両方からプライバシの問題に取り組み，両者の需要を満たす要件を調査する必要がある．

ガイドラインの提案だけではなく，サービスやアプリケーションとしてユーザのプライバシを保護することが必要である．確かにガイドラインは有効ではあるが，それだけではデジタル通信時代のプライバシのあり方として世の中で機能させることは容易ではない．ユーザが自身のプライバシに関わる情報を管理できる社会の実現のためにも，どこまでの情報を取得するとユーザのプラバシを脅かすのかという判断をするシステムを構築，普及する必要がある．そのような，判断基準となるシステムを作成するために，より多くのプライバシに関する調査や，複数の情報を組み合せるアルゴリズムの提案をによって，プライバシ情報を判断するシステムを構築することを通して，今後のプライバシのあり方の指針を示すことが今後の課題である．

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] Inc Amazon.com. Amazon.co.jp: 通販 -ファション、家具から家電まで. http://www.amazon.co.jp, 12 2009.

[2] NAVITIME JAPAN. 地図検索 | navitime. http://www.navitime.co.jp/, 12 2009.

[3] 総務省行政管理局. 個人情報の保護に関する法律. http://law.e-gov.go.jp/htmldata/H15/H15HO057.html, 12 2009.

[4] Alan Westin. *Privacy and Freedom*. New Jork Atheneum, 1967.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

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

[16] 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.

[17] T. KARAGIANNIS. Blinc : Multilevel traffic classification in the dark. *ACM Sigcomm, Philadelphia, PA, Aug. 2005*, 2005.

[18] M. Zalewski and OS Passive. Fingerprinting tool. *http://lcamtuf.coredump.cx/p0f.shtml*, 1 2010.

[19] Inc mixi. [mixi]. `http://mixi.jp`, 9 2009.

[20] Twitter. twitter. `http://twitter.com`, 9 2009.

[21] 高木浩光. Bluetooth で山手線の乗車パターンを追跡してみた. http://takagi-hiromitsu.jp/diary/20090301.html, 2009. 12.

[22] Evernote Corporation. Welcome to your notable world — evernote corporation. http://www.evernote.com/, 1 2010.

[23] Apple. Apple - download music and more with itunes. play it all on ipod. http://www.apple.com/itunes/, 12 2009.

[24] OECD RECOMMENDATION CONCERNING AND GUIDELINES GOVERNING THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA. O.E.C.D. Document C(80)58(Final), 10 1980.