

卒業論文 2009年度（平成21年度）

デジタル情報収集による
ユーザ追跡のリスク分析と対策の提案

慶應義塾大学 総合政策学部

氏名：上原 雄貴

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22年2月18日

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、ユーザプライバシー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ー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、ユーザプライバシーを守るために、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、ユーザプライバシー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、ユーザプライバシー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ーに対する影響を考察した。そして、ユーザのプライバシーを保護するために、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

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 information 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

第1章 序論	1
1.1 ユーザが発信する情報とその利用	1
1.2 本研究の目的	2
1.3 本論文の構成	3
第2章 デジタル情報とプライバシー	4
2.1 デジタル情報収集の現状	4
2.1.1 ユーザに発行した識別要素に基づいて情報を収集する手法	4
2.1.2 ユーザ自身がデータを直接相手に送信する手法	5
2.1.3 ユーザが発信する情報を受動的に取得する手法	6
2.2 プライバシの脅威となる情報	7
2.3 ユーザのプライバシーの考慮	9
2.4 情報収集者が取得できる情報	10
2.4.1 同じネットワークに接続していないユーザ	10
2.4.2 サービス提供者	12
2.4.3 同一セグメントに存在するユーザ	14
2.4.4 ネットワークを管理するユーザ	15
2.5 複数の情報統合によるリスク	16
2.6 本論文の着眼点	17
2.7 まとめ	17
第3章 関連研究	18
3.1 ソーシャルネットワークを利用した情報収集	18
3.2 Web 上での情報収集	18
3.3 ベイズ統計を用いたユーザ嗜好の分析	19
3.4 ブラウザ情報を利用した個人識別	20
3.5 情報統合に対する対策の検討	20
3.6 まとめ	20
第4章 デジタル情報を用いたユーザ特定手法	21
4.1 ネットワーク管理者と取得情報	21
4.1.1 前提	21
4.1.2 パケットのヘッダ情報	23

4.1.3	ホスト識別による調査	23
4.2	同一セグメント上のユーザと取得情報	33
4.2.1	前提	34
4.2.2	共有ホスト名	34
4.3	第三者であるユーザと取得情報	35
4.3.1	前提	36
4.3.2	Bluetooth	36
4.4	検証情報統合によるリスク	37
4.5	まとめ	37
第5章	検証	38
5.1	パケットのヘッダ情報	38
5.1.1	検証手法	38
5.1.2	ホスト識別システム	38
5.1.3	設計概要	39
5.1.4	ホスト識別に用いる情報	40
5.1.5	検証環境	41
5.1.6	検証結果	41
5.1.7	考察	44
5.2	共有ホスト名	44
5.2.1	検証手法	44
5.2.2	設計概要	45
5.2.3	検証環境	45
5.2.4	検証結果	45
5.2.5	考察	47
5.3	Bluetooth デバイス名	47
5.3.1	検証手法	47
5.3.2	設計概要	47
5.3.3	検証環境	48
5.3.4	検証結果	48
5.3.5	考察	49
5.4	検証した情報の統合	50
5.5	まとめ	51
第6章	ガイドラインの提案	52
6.1	一般ユーザのガイドライン	52
6.2	開発者、管理者のガイドライン	52
6.3	ガイドラインの充足度の検討	54
6.4	まとめ	56

第7章 結論	57
7.1 まとめ	57
7.2 今後の展望	58
謝辞	59

目次

2.1	クライアントがサーバから付与された識別要素を管理	5
2.2	ユーザがサーバに情報を直接送信	6
2.3	ユーザが全体に情報を発信し、受信側がデータを処理	7
2.4	プライバシマークサンプル	9
2.5	TRUSTe マークサンプル	9
2.6	Bluetooth の概要	11
2.7	Bluesoleil	12
2.8	Apache log	13
2.9	Mail log	14
2.10	アカウントを用いたユーザプロフィール作成の概要図	16
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	19
4.1	ユーザ特定手法の全体図	22
4.2	パケットヘッダ情報の収集システムの概要	22
4.3	ホスト A の送信先 IP アドレス・ポート番号	24
4.4	ユーザ B の送信先 IP アドレス・ポート番号	24
4.5	MacOSX と WindowsXP の利用送信元ポート番号	25
4.6	MacOSX の起動時のパケットの発信タイミング	27
4.7	Windows Vista の起動時のパケットの発信タイミング	27
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	28
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	28
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	30
4.11	時間におけるホストの IP アドレス遷移回数	32
4.12	パケットヘッダ情報の収集	34
4.13	MacOSX の Finder	35
4.14	Bluetooth 情報の収集	37
5.1	システムの動作概要	39
5.2	ホスト識別手法の設計	40
5.3	WIDE 合宿ネットワークトポロジ図	42
5.4	共有ホスト名を用いた実験のネットワーク概要図	46
5.5	共有ホスト名を用いたユーザ生活モデル	46
5.6	Bluetooth デバイスの検出	48

5.7 Bluetooth デバイスの検出によるユーザのライフタイム	50
6.1 一般ユーザのガイドライン	53
6.2 開発者, ネットワーク管理者のガイドライン	53
6.3 OECD8 原則	55

表 目 次

2.1	ペイロード解析によるユーザ識別結果	16
4.1	OS と利用発信元ポート	26
4.2	識別要素とする対象ツール一覧	31
4.3	送信先 IP アドレス上位リストの類似調査	33
5.1	WIDE 合宿ネットワークにおける実験結果	42
5.2	共有ホスト名の手法検証の実装環境	45
5.3	Bluetooth デバイスアドレス検証手法の実装環境	47
5.4	Bluetooth デバイスアドレス検証結果	48
5.5	検証で利用した情報	50
6.1	ガイドラインの充足度	56

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、ユーザプライバシー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ー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ー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ーを保護する必要がある。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ー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つて、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、プライバシーとは Privacy and Freedom[4]で述べられている、“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ー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ィー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ーが脅かされつつある。

現在、ユーザのプライバシー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ー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ー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ー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ー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ー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ー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ー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、本論文に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ー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によって、自身のプライバシーを守らなければならない。プライバシーの保護と、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ー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ー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ー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ー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、プライバシー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ー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ー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、プライバシー守るためには、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ー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 デジタル情報とプライバシー

本章では、デジタル情報の収集の現状を述べ、3つの手法に分類するとともに、観測者ごとに取得できる情報について述べた。本論文はホストが発信する情報に着目することを示した。

2.1 デジタル情報収集の現状

今日、デジタル技術の発展によって、情報の収集方法は多岐に渡るようになった。これによって、サービス側はユーザに応じたサービスを提供することで、ユーザはより質の高いサービスを受けることができる。しかし、同時に個人と関係のある情報がネットワークでやり取りされるようになり、ユーザのプライバシーは脅かされている可能性がある。そこで本論文では、ユーザの情報はどのようにやり取りされるのかという現状を3つのモデルに分類する。また、デジタル情報を収集するにあたり、ユーザにとって様々な利点と欠点が発生する。本章ではその技術の例を具体的に挙げる。

2.1.1 ユーザに発行した識別要素に基づいて情報を収集する手法

デジタル情報を収集する上で、ユーザを判別する手法に、識別要素を判別対象とするユーザに付与する手法がある。これは、サーバ側はコンテンツの情報やユーザ情報を識別要素として発行し、ユーザ自身に保持させる仕組みである。これを利用することによって、ユーザに応じた個別のサービスが可能となる。どのようにそれが実行されるかを図2.1に示す。

まず、サーバはクライアントが接続をしてきた場合に識別要素を発行する。この識別要素に含まれている情報はユーザを識別するためのIDや、サービス設定を保存するものなどである。識別要素を保持しているユーザが再度そのサーバを利用する際に、自身が保有する識別要素をサーバに送信する。これによって、サーバは識別要素として示される情報をもとに、ユーザに応じた情報を送信する。

ユーザは、識別要素を自身で管理するが、識別要素を利用していることを理解した上で容易に管理できる場合と、本人が知り得ぬところで識別要素を付与され、管理が難しい場合がある。例えば、HTTP Cookieは多くのサイトで利用され、Webブラウザの設定でCookieを制限できることが容易である。最近では、Firefox[5]などのブラウザは設定によって、HTTP Cookieや履歴など痕跡を残さずにWebアクセスが可能となるため、HTTP Cookieの利用に関してユーザのプライバシーに配慮がされていると言える。

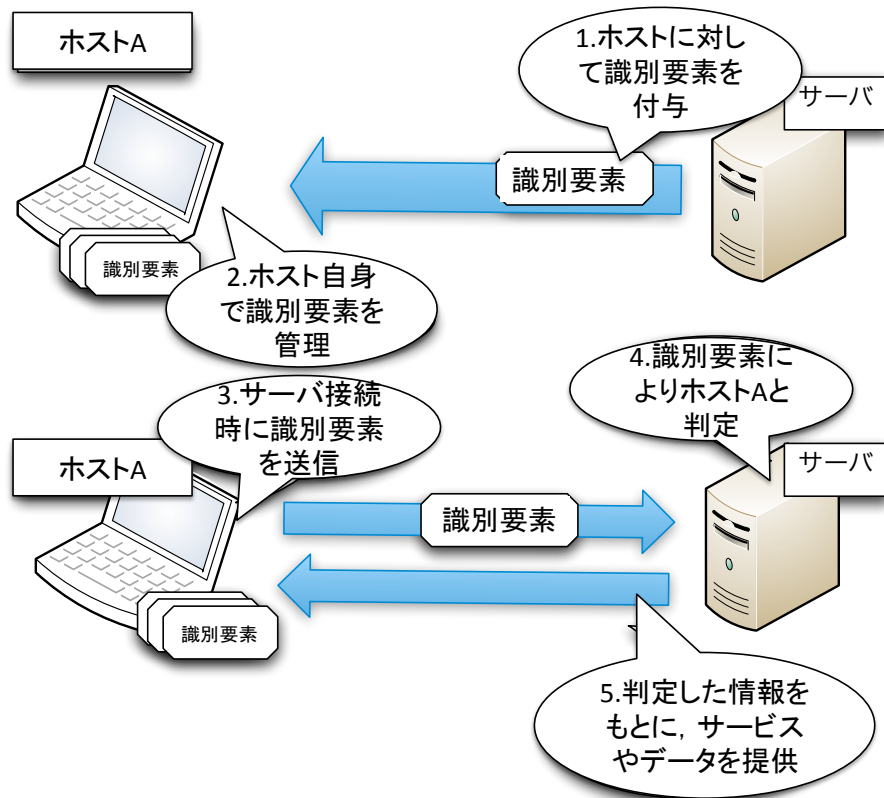


図 2.1: クライアントがサーバから付与された識別要素を管理

HTTP Cookie とは別に、本人が知り得ぬところで識別要素を付与され、管理が難しいという識別要素の例に Flash Cookie が挙げられる。この Flash Cookie は、Web ブラウザだけでは削除することはできない。保存できる情報量は、HTTP Cookie が 4KB に対して、100KB とより多くの情報が記録できる。Flash Cookie を確認・管理する方法も、Adobe[6] の Flash Cookie のページで設定しなければならない。また、HTTP Cookie に比べて認知度が低いため、存在を知らないユーザも多い。Flash Cookie の危険性の調査がアメリカで行われた [7]。この調査では QuantCast[8] に掲載されているユーザ利用率が高い上位 100 サイトのうち、54 のサイトで Flash Cookie が利用されている。そのため、ユーザが管理しにくい、もしくは知りえぬ識別要素がネットワーク上でやり取りされていることを示している。このように Flash Cookie はユーザが知り得ぬところで識別要素が付与され、かつユーザ自身が識別要素の管理が困難である。そのため、プライバシーを脅かす可能性がある。

2.1.2 ユーザ自身がデータを直接相手に送信する手法

ユーザが自身のホストにアプリケーションをインストールする手法や、情報をサーバに送る手法などが挙げられる。そのモデルを図 2.2 に示す。ユーザが同意した上で、アプリケーションのインストールを行い、サービスにログインした状態でサイトを閲覧などをす

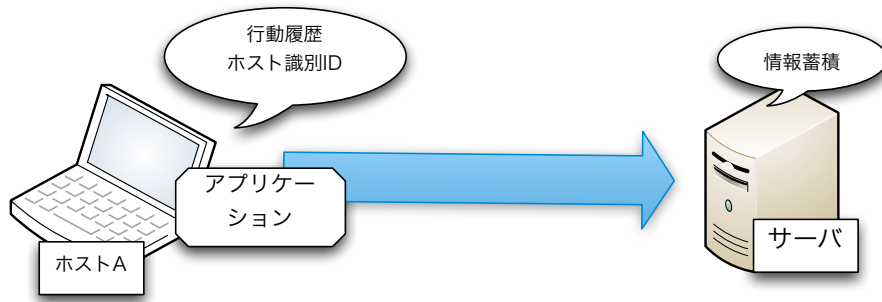


図 2.2: ユーザがサーバに情報を直接送信

ることで、サーバに情報を送信する。品質向上のためにバグレポートや利用履歴を送信するなどの機能を備えたアプリケーションもその一例である。また、Web サイト閲覧時にログインをした状態で Web ページにアクセスすると、おすすめの情報を提供するレコメンドサービスなども Amazon をはじめとして多く存在する。

既存研究としては、使用するホストなどの端末にエージェントをインストールすることによって、ユーザの傾向や振る舞いを識別する手法などがある。このアプリケーションによる研究は、高度なパーソナライズ実現のためのユーザプロファイル統合サービスエージェントの設計 [9] をはじめとして広く研究されている。他にも、インターネット上でのトレンドを知る調査にもこのような手法は利用されている。ネット視聴率白書 2008-2009 [10] でもこの手法が採用されている。しかし、これらの手法はクライアントすべてにアプリケーションを導入しなければならないため、情報を取得した側のコストが高い。また、アカウント名やパスワードを利用して、サービスにログインする場合もこのモデルに含まれる。サービスに同意を得た上でアカウントを作成し、ログインして情報を発信するためである。

2.1.3 ユーザが発信する情報を受動的に取得する手法

ユーザが通信をする際にネットワーク全体や周囲に対して情報をブロードキャストやマルチキャストによって通知する場合がある。また、Bluetooth や RFID など機器を利用して、情報を発信する方法もある。それらの状態を取得して、サービスを提供したり、トラフィック情報を収集することができる。また、ネットワークにおいて通信のやりとりからアプリケーションを推測するなど、データ収集側は受動的に個人情報を収集する。そのモデルを図 2.3 に示す。

トラフィックを分析する事によって、個人情報を取得することができる Deep Packet Inspection (DPI) と呼ばれる技術がある。この技術はリアルタイムにフローパターンやパケットの特徴や振る舞い、パケットの制御情報などからトラフィックを解析する技術である。この技術を利用して行動分析型ターゲット広告をする企業に NebuAd [11] が挙げられる。NebuAd はユーザの同意を得てサービスを提供する代わりに、ユーザの行動を分析して広告するというサービスを提供している。取得した情報をデータベース化し、その統計から

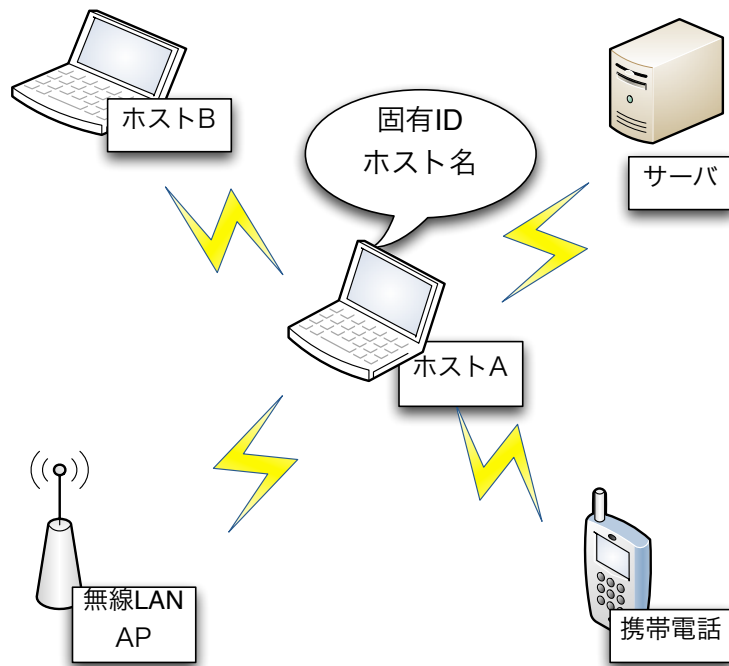


図 2.3: ユーザが全体に情報を発信し，受信側がデータを処理

アプリケーションの特定を行う。日本国内でも、ISP が P2P プロトコルを利用しているホストのトラフィック制御にこの技術は利用している。しかし、この技術は通信の秘密に抵触する可能性があるというデメリットを保持している。

受動的にトラフィックを収集、解析する研究は多く行われている。例えば、ネットワークトラフィックに着目することでセキュリティインシデントをリアルタイムで発見する研究に *Passive Network Discovery for Real Time Situation Awareness*[12] などがある。この研究は様々な *Passive finger printing* を利用することによって、ネットワークに負荷をかけることなく情報を取得し、ユーザを特定する。この研究で用いている *Passive finger printing* によって取得できる情報は稼働中のホストや OS 情報、ホストの役割、提供サービス、プロトコル、ネットワークの IP アドレス設定である。他にも、*BLINC : Multilevel traffic classification in the dark*[13] はパケットのヘッダ情報を利用してネットワークに流れるプロトコルやアプリケーションを把握する研究など様々である。

このように、デジタル情報を収集する上で様々な技術がある。その反面、これらの技術を利用することで、ユーザのプライバシーは脅かされる可能性がある。プライバシーを脅かされる情報はどのようなものがあるのかを述べる。

2.2 プライバシの脅威となる情報

ユーザにとってネットワーク上でプライバシーの脅威となる情報について述べる。ネットワークの構成やレベルにおいて、ユーザやその通信内容の匿名性はある程度保持される。

しかし、情報によってはネットワーク上でユーザを判別・追跡することができる。ここで述べる情報とはユーザを特定できる識別要素であり、ユーザが利用している機器や情報が一意に特定できる情報を指す。

ネットワーク上でのユーザを一意に特定できる識別要素は多くあるが、ユーザが利用する情報だけではない。以下に具体例を数点紹介する。

- クレジットカード情報

クレジットカードは、商品購入やサービスを受ける際の決済方法に利用される。クレジットカードには契約番号や名前、セキュリティコードなどが記載されている。契約番号は固有であり、ユーザを分ける識別要素となる。クレジットカード決済を利用できるサービスや店は多く、ユーザの購買履歴を閲覧することで、クレジットカードを保有しているユーザの好みの傾向や行動を追跡することが可能となる。

- 携帯電話の固有 ID

携帯電話の固有 ID を利用したサービスにかんたんログインというものがある。携帯電話の固有 ID で容易にユーザの認証をすることができる。その反面、一意に情報が知られてしまうと、変更が難しいため、悪用される危険がある。

- 自動車ナンバ

自動車のナンバを収集し、犯罪捜査や証拠として扱うシステムが自動車ナンバ自動読み取り装置 (Nシステム) である。これは、警視庁が各拠点に設置している。車のナンバは同じものがない一意の識別要素であるため、車のナンバを利用することで、その持主の移動情報を取得することができる。

- IC カード

PASMO[14] や Suica[15] に代表されるように、IC カードの利用機会が増加している。IC カードには固有の ID が割り当てられており、ID からユーザを識別することができる。また、交通機関や買い物の精算をした場合、IC カードに記録が残るため、ユーザの行動履歴や場所情報、場合によっては住所を取得することができる。

- ホスト情報

ホストに関する情報は個人情報となりやすい。一意にホストを特定することができる MAC アドレスや、ホストの名前などホストに関する情報が挙げられる。他にも、ホストから発信する情報から、ユーザの行動履歴や傾向を分析することができる。

- サービスアカウント名

Google[16] をはじめとするサービスはアカウント名を付与されてログインすることでサービスを受けることができる。アカウント登録の際には個人情報を入力する場合がある。ログイン時のユーザの行動履歴を追うことで、よりユーザに応じたサービスが可能になる。また、個人情報を入力することや、Web ページでのアクセス履歴など、アカウント情報が多くのユーザのプライバシーに関する情報を含む場合がある。

2.3 ユーザのプライバシーの考慮

近年、個人情報に関する事件が多発したため、サービス、コンテンツ提供者もユーザのプライバシーについて対策する必要がある。プライバシーの起源は、Samuel D. Warren, Louis D. Brandeis らが 1980 年に発表した The Right to Privacy [17] にある the right to be let alone であると言われている。そこからプライバシーは個人に関する情報が本人に同意を得ることなく利用されないことを指す。近年は情報化に伴ない、様々なプライバシーを守る法律や、技術が確立されている。

このような情報化の背景により、ユーザも利用する企業やサイトがユーザプライバシーを守るかという点に注意を払うようになった。サービスやサイトがプライバシー対策を施しているかどうかの判断基準の一つにプライバシーマーク制度 [18] がある。これは日本工業規格 JIS Q 15001 個人情報保護マネジメントシステムの要求事項に適合して、個人情報について適切な保護措置を講ずる体制を整備している事業者等を認定する制度であり、プライバシーマークをサイトに記載している。図 2.4 にプライバシーマークのサンプルを示す。



図 2.4: プライバシマークサンプル



図 2.5: TRUSTe マークサンプル

このような認証マークをほどこす事例は数多くあり、TRUSTe もその一つである。TRUSTe は情報の完全公開と利用者の同意を基本理念としており、1980 年 9 月 23 日に経済協力開発機構の理事会で採択された「プライバシー保護と個人データの国際流通についての勧告」の中に記述されている、8つの原則をもとにしている。図 2.5 にそのサンプルマークを示す。これらのマークを付与している所では、個人情報の利用規約を記載し、ユーザに同意を得るケースが多い。これらのマークをサイトに載せることによって、利用者はそのサイトを信頼できるかの基準とすることができる。

企業だけではなく、通信事業者にもユーザのプライバシーを保護する法律が存在する。電気通信事業法第 4 条第 1 項で通信の秘密において“電気通信事業者の取扱中に係る通信の秘密は、侵してはならない”として定めている。

通信の秘密が焦点となった例に、NTT のぷらら [19] が Winny 規制サービス提供が通信の秘密に抵触する恐れがあることが総務省から通達されるという出来事があった。これは、NTT のぷららが Winny の利用を禁止するために、ユーザの通信を閲覧し、規制するというものである。ここで焦点になったのはユーザの同意を得ていないという点である。この通達を受け、ユーザの同意を得るという条件の下で Winny の規制サービス [20] に修正した。日本国外でもこのような事例が多くある。前述した米 NambuAd では、ユーザの

同意を得るオプトアウト形式でユーザの通信を取得して広告を出すというサービスを展開していたが、多くの提携先やユーザが難色を示し、最終的にはCEOが辞任しアメリカから撤退するという事例があった。これらの事例から、個人情報を収集・解析をする際には、ユーザの同意や理解を得ることが必要と言える。

また、個人情報を所得管理する側が利用する技術は多岐に渡る。例えば、ユーザの通信をすべて暗号化することや、データの集計して統計にする際にランダムに変数を変化させるk-匿名技術 [21] などがあげられる。これによって、ユーザのプライバシーを侵害しないまま、統計を取ることが可能になった。

他にもユーザのプライバシーを保護する法律に電波法が挙げられる。電波法にも秘密の保護として第五十九条に”何人も法律に別段の定めがある場合を除くほか、特定の相手方に対して行われる無線通信（電気通信事業法第四条第一項又は第九十条第二項の通信たるものを除く。第百九条において同じ。）を傍受してその存在若しくは内容を漏らし、又はこれを窃用してはならない。”[22] と明記されている。つまり、管理者や同一セグメントに存在するユーザが、他のユーザのパケットを取得した場合、同意を得た場合を除き、そこから判明した情報を漏らしてはいけない。そのため、管理者はトラフィックから得た情報を秘匿する必要がある。

2.4 情報収集者が取得できる情報

ユーザのプライバシーを守るために多くの規定や技術が提案されている。しかし、ユーザも自身のプライバシーを侵害するような情報はどのようなものがあるかを知る必要がある。どのような情報を発信しているかを知ること、守らなければならない情報を明確にすることでよりプライバシーの脅威を低減できるからである。

本論文では、インターネットにおけるユーザのプライバシーに焦点を当てて述べる。その理由は、インターネットによって多くのユーザが時間場所を問わずに情報のやりとりができるため、ホスト情報が利用者の個人情報となったためである。日本におけるインターネット人口は1997年では572万人から、10年後の2007年には8227万人 [23] と爆発的に増加し、更に増え続けている。ここから、日本人の大半の情報はネットワーク上から得る可能性があることを示している。

ユーザのプライバシーが脅かされる情報が取得されるかどうかは、情報収集者と情報収集対象者とのネットワークの関係によって異なる。そこで、ネットワークにおける第三者、同一セグメントであるユーザ、ネットワークにおける管理者の3パターンに分類して述べる。

2.4.1 同じネットワークに接続していないユーザ

同じネットワークに接続していないユーザが取得できる情報について述べる。

実際に、悪意のあるユーザがターゲットとするユーザと同一セグメント上に存在しない場合を考察する。この場合、悪意あるユーザがとれる行動は非常に制限されるが、主な手

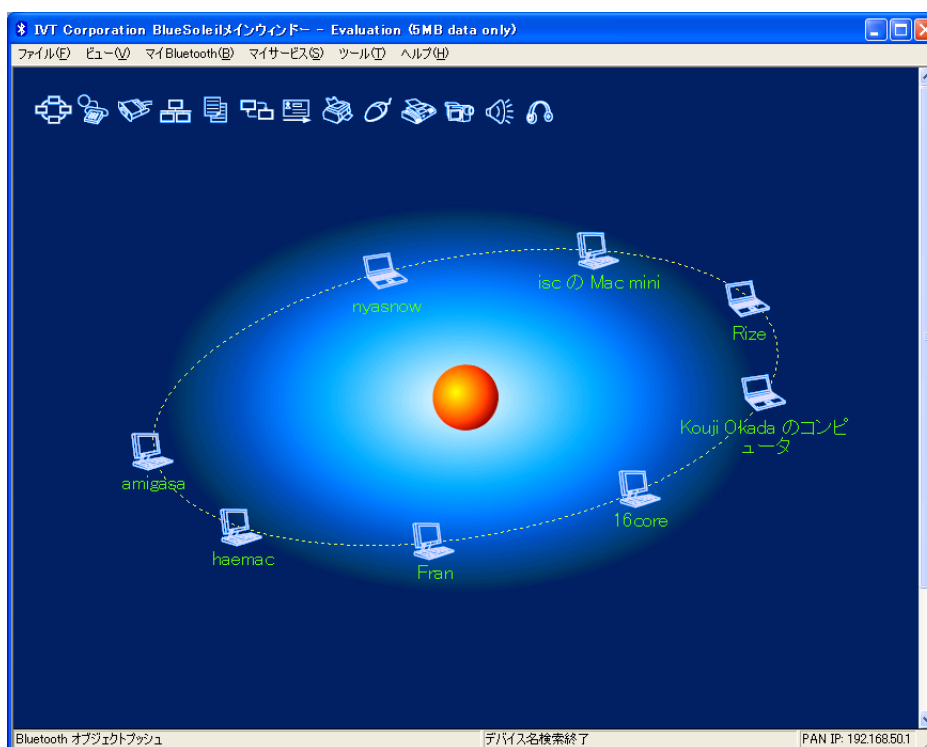


図 2.7: Bluesoleil

Bluetooth デバイスアドレスの探索では BlueSoleil[25] というツールが存在する。このツールは、図 2.7 のように Bluetooth を探索し、視覚的に表示するツールである。

- RFID

Radio Frequency Identification(RFID) は無線タグにより人やモノを識別、管理する技術全般を指す。タグ自身から数 cm から数 m の電波を発しているものから、自身から電波を発しないパッシブタグと呼ばれる IC タグなど幅広く存在する。RFID を利用することで、ユーザの位置情報やものの場所の把握・管理を促進することができる。

- Web 検索, SNS

Web 検索や Social Network Service を利用してユーザの実世界における人間関係や行動を把握することができる。特に、SNS サイトでは写真やプロフィールを公開している場合が多々ある。それだけではなく、本人に関わる記事が Wikipedia といったサイトに記載される場合があり、詳細な情報を得ることができる。

2.4.2 サービス提供者

サービス・コンテンツ管理者が利用できる情報を挙げる。

- Cookie

Third-party Cookie を利用することでユーザを追跡することが可能となる。その

```

203.178.128.72 - - [10/Nov/2009:20:00:51 +0900] "GET / HTTP/1.1" 200 44 "-"
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; ja-JP-mac; rv:1.9.1.5)
Gecko/20091102 Firefox/3.5 .5 Jingoo/0.1.0"
203.178.128.72 - - [10/Nov/2009:20:00:51 +0900] "GET /favicon.ico HTTP/1.1" 404
209 "-" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; ja-JP-mac; rv:1.9.1.5)
Gecko/20091102 Firefox/3.5.5 Jingoo/0.1.0"
203.178.128.72 - - [10/Nov/2009:20:00:54 +0900]
"GET /favicon.ico HTTP/1.1" 404 209 "-" "Mozilla/5.0 (Macintosh; U; Intel Mac
OS X 10.6; ja-JP-mac; rv:1.9.1.5) Gecko/20091102 Firefox/3.5.5 Jingoo/0.1.0"
203.178.128.72 - - [10/Nov/2009:20:04:53 +0900] "GET /nakajima/grah HTTP/1.1"
301 251 "-" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; ja-JP-mac; rv:1.9.1.5)
Gecko/20091 102 Firefox/3.5.5 Jingoo/0.1.0"

```

図 2.8: Apache log

手法の一つとして、Webサーバを管理し、そこにアクセスするユーザのトラッキングが挙げられる。前述した Cookie はユーザに即したデータを取得できるという利点があるが、逆にその Cookie を利用することによって、ユーザの行動履歴を把握することが可能となる。最近のサービス・コンテンツ提供者は、サーバを分散し、複数のドメイン間でデータのやり取りをしている場合がある。これを利用することでユーザのトラッキングが可能となる。

- Webサーバのログ

Webサーバを設置することで、アクセスしたユーザのログを取得し、ユーザが利用するホストの情報を利用することが可能である。図 2.8 の例で挙げられるように、Apache の log を解析することでアクセスしたホストの日時、ブラウザ名、アクセス元の国、OS を判別することができる。他にも Cookie を利用している場合は Cookie の値が出力されるため、ユーザがどのように Web を一覧しているのか追跡することが可能である。

- Mail のログ

受信した Mail のヘッダ情報や Mail サーバのログを閲覧することでホストを推測する。図 2.9 のように、利用ユーザのメールアドレスや、送信元のメールアドレスや転送している場合は、転送しているメールサーバとアドレスの情報がログとして残る。

- GPS サービス

Global Positioning System(GPS) は元は軍事用に開発されていたが、民間に開放されることで、GPS に関するサービスが施行された。近年は、GPS 付き携帯電話が増えることで、容易に人の位置情報を得ることができるようになった。GPS を利用

```
Dec 20 00:00:40 mail postfix/local[4191]: 775544C138: to=jnakajima@sfc.wide.ad.jp,
relay=local, delay=0.04, delays=0.02/0/0/0.02, dsn=2.0.0, status=sent (delivered to
maildir)
Dec 20 00:00:40 mail postfix/local[4191]: 775544C138: to=jnakajima@sfc.wide.ad.jp,
relay=local, delay=0.05, delays=0.02/0/0/0.03, dsn=2.0.0, status=sent (forwarded
as 7FC794 C536)
Dec 20 00:00:41 mail postfix/smtp[4800]: 7FC794C536:to=nakajimapanda@gmail.com,
orig_to=nakajima@sfc.wide.ad.jp, relay=gmail-smtp-in.l.google.com[209.85.222.47]:25,
delay=1.5, delays=0.01/0/0.61/0.85, dsn=2.0.0, status=sent (250 2.0.0 OK
1261234841 13si1642044pzk.59)
```

図 2.9: Mail log

するサービスは多数あり、NAVITIME などの道案内など多岐に渡る。容易に GPS を利用できるようになったが、GPS 機器を保持したユーザを追跡が可能になるなどプライバシーと密接な関わりがある。

- SaaS

Software as a Service(SaaS) というサービスの提供者は、契約したユーザが利用したソフトウェアとその挙動を取得することができる。SaaS は契約したユーザが必要な機能のみを選択してソフトウェア提供を受けるサービスである。契約内容やソフトウェアによってプライバシーに関する情報は変化するが、一般的には SaaS 管理者は提供するソフトウェアにおいて個人情報を取得・管理することができる。

- PaaS

Platform as a Service(PaaS) は SaaS を拡張し、プラットフォームや開発環境全般を提供するものである。SaaS に開発環境なども追加しているため、ソフトウェアに限らないため、SaaS よりも情報全体を管理することができる。

2.4.3 同一セグメントに存在するユーザ

同一セグメント上に存在するユーザはどのくらい情報を取得できるか考察する。スイッチの技術向上によって、宛先と関係のないポートにはパケットを送信しないようになった。しかし、ホストは同一セグメントにおいてブロードキャスト、マルチキャストなどで様々な情報を送信している。同一セグメントにおいてどのような情報が取得可能であるかを以下に挙げる。

- ARP

Address Resolution Protocol(ARP) は、イーサネット環境において、IP アドレスから MAC アドレスを得るために用いられるプロトコルである。ARP を利用するこ

とによって、ネットワークの構成や同じセグメントにいるホストの MAC アドレスや IP アドレスを取得することができる。しかし、スイッチなどネットワークの構成で同じスイッチ間でしか情報を取得できない場合がある。

- mDNS

Multicast DNS(mDNS) は DNS の代わりに、サービスに関する情報を提供する。mDNS はサービスの種類、ホスト名、IP アドレスやポート番号の情報が含まれる。主に MacOSX で利用される傾向にあり、Bonjour などファイル共有時に利用される傾向にある。

- NetBIOS

SMB などファイル共有のアプリケーションに利用されるネットワークサービスを呼び出すための API の規約。MS-DOS 環境で多く使われている。NetBIOS 名はユーザやアプリケーションが自由に決定することが可能である。SMB などのファイル共有のアプリケーションはホスト名を NetBIOS 名に利用しているケースが多々ある。

- DHCP

ホストは IP アドレスを取得するために、Dynamic Host Configuration Protocol(DHCP) を利用する。その際に、DHCP サーバを検索するためにネットワーク全体にブロードキャストをする。もし、DHCP サーバ探索をしているホストの MAC アドレスを事前に入手しておけば、そのホストがいつネットワークに参加し、離脱したかなどの情報を得ることができる。

2.4.4 ネットワークを管理するユーザ

ネットワーク管理者はどのくらいの情報を取得できるかを考察する。管理者はネットワークのポリシーや構成に依存するが、非常に多くの情報を取得することができる。本論文ではネットワーク管理者をネットワークポリシーを自由に設定できる権限を保有するユーザと定義する。そのためネットワーク上にあるすべてのパケットのペイロードを閲覧可能であると言える。

筆者は、ペイロードを閲覧することで、精度の高いプロファイルを作成できるユーザ識別システム [26] を実装した。このシステムは管理者が動的に個人情報を取得・管理することにより、インシデント発生時の対応を支援し、ユーザの利用や挙動に関する情報を IP アドレスとひも付け IP アドレスを基準にしたユーザのプロファイル作成している。

このシステムで IP アドレスと結びつける情報はサービスのアカウント名を対象としている。対象とするアプリケーションの ID は Google アカウント、Yahoo アカウント、MSN メッセンジャで使用するアカウント情報(メールアドレス)、Mixi アカウント ID、Twitter アカウント ID の 5 つとした。5 つのアカウント名を識別要素としてユーザのプロファイルを作成し、ホストの所有者を特定する。そのため、IP アドレスが変化しても同じアカ

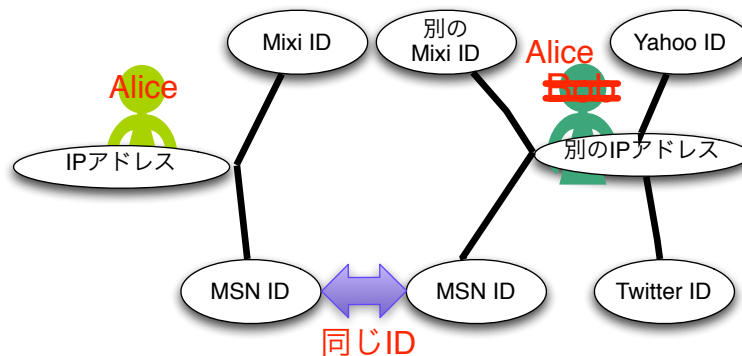


図 2.10: アカウントを用いたユーザプロフィール作成の概要図

表 2.1: ペイロード解析によるユーザ識別結果

全ホスト数	識別したホスト数	検出割合
128	81	63%

アカウント名を利用しているとシステムが判断すると一人のユーザに統合される。図 2.10 にアカウントを用いたユーザプロフィール作成の概要図を示す。

このようなモデルで、トラフィックから個人情報を収集し、実際に筆者の所属する研究室で検証した。研究室には 100 人程度のユーザが存在し、ホストは 200 台程度存在する。研究室のネットワークの上流でミラーリングをし、すべてのトラフィックを収集した。収集した期間は 2009 年 1 月 21 日 15 時 57 分から 1 月 22 日 15 時 56 分である。

実験結果を表 2.1 に示す。ユーザのプロファイルできた IP アドレス数を、取得した IP アドレス総数で割ることによって導き出した。評価実験の結果により、128 台のホストのうち 81 台のホストのユーザを識別することができた。これはすべてのホストの 63% を識別している。筆者の研究室では常に起動しているサーバが多数あるため、128 台中 81 台までのホストを識別できたため、有効であると言える。

このことから、利用サービスのアカウント名を取得するだけで、ネットワークにおいて、ユーザの識別はホストを複数利用していても可能であり、プライバシーの脅威となる。

2.5 複数の情報統合によるリスク

前述した手法やユーザの同意を得ることでユーザのプライバシーを脅かさずに、もしくは同意の上で情報を取得することができるようになった反面、個々の情報だけ保護することの意味が無い場合が多々ある。例えば、保護した情報を複数集めることによってプライバシーに脅威を与える場合や、個人情報に値しない情報を複数提供することによって個人がトラッキングされるケースが多々ある。そのため、どのような情報を統合されるとユーザの

プライバシーが脅かされるかを認識しなければならない。

複数の個人情報統合される場合、どのようなことが判別できるのか。前述した識別要素を組み合わせることで、よりユーザのプロファイルの作成が可能となる。例えば、クレジットカードの利用履歴とそのクレジットカードを利用したホストの MAC アドレスを利用することで、特定のユーザの場所や通信をトラッキングできる。他にも、トラッキングを目標としているユーザの MAC アドレスを保持している場合、Apache のログを組み合わせることでホストの利用している Web ブラウザや OS などの情報を結びつけてプロファイルを作成される危険がある。

2.6 本論文の着眼点

本論文は、第 2.1.3 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 2.2 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ー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ー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ーを脅かす可能性があるかを検討する。

2.7 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ー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ー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[27] 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ー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、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ます、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ーの脅威について理解せずに、情報を書きこむことに対する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 2.1 と図 2.2 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 [28] 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

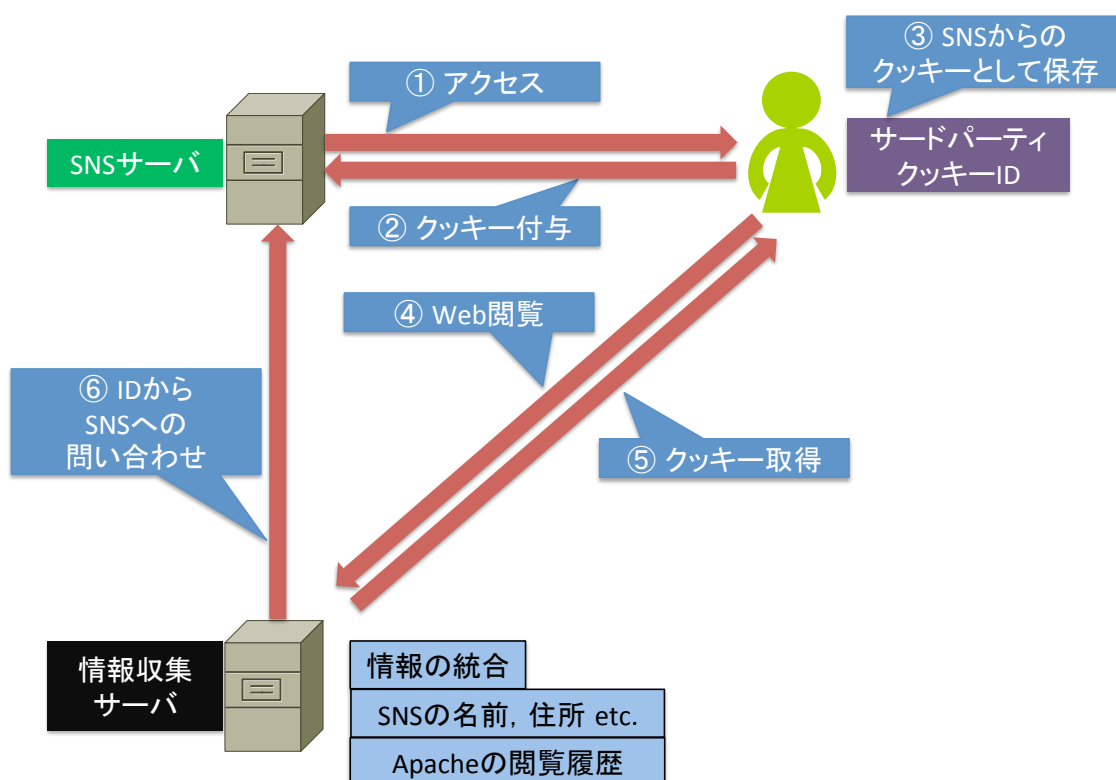


図 3.1: Cookie を利用して得た SNS 情報と Apache ログの組み合わせ手法

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 2.2 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[29] 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る、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 2.2 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 [30]. 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 2.2 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ー侵害とその対策 [31] 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ー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 [32][33]. 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・プライバシーの保護 [34] 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ー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, プライバシー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る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ー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ーを考慮しながらも、ユーザの情報を収集する必要があるため、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ー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ィールを作成する。最後に、同一ネットワークに接続していないユーザは、Bluetoothを利用したプロフィールを作成する。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ー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ー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ー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。

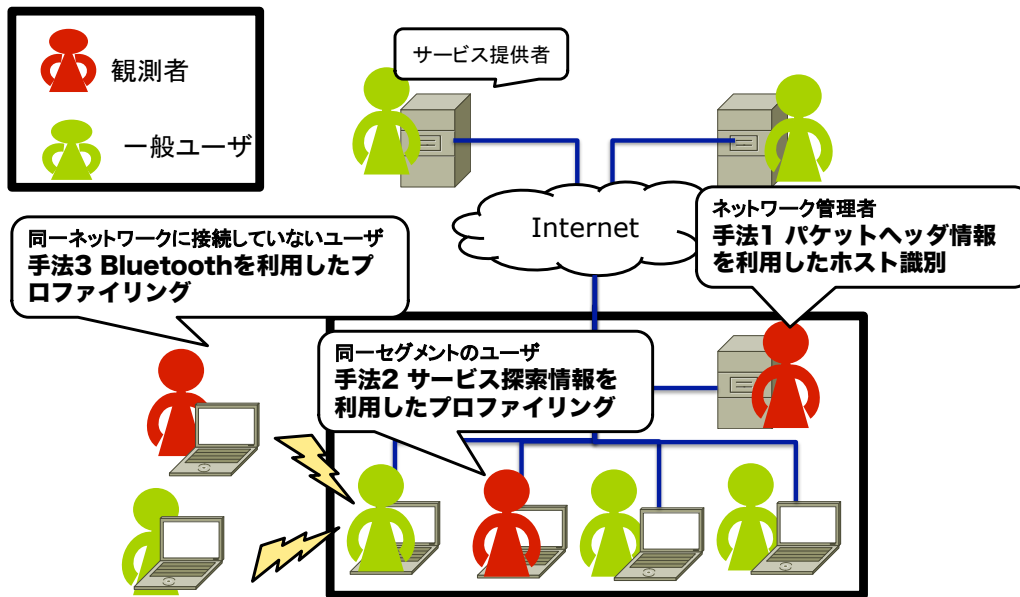


図 4.1: ユーザ特定手法の全体図

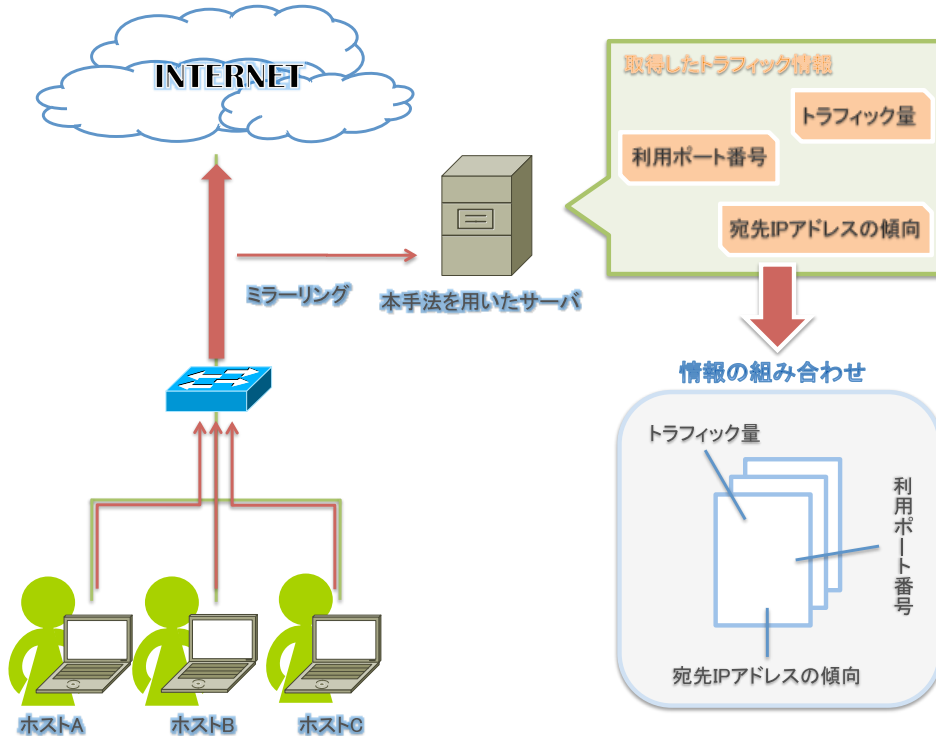


図 4.2: パケットヘッダ情報の収集システムの概要

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 [13]。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ーを脅かさないと断言できない。

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 Fingerprinting [35] 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。

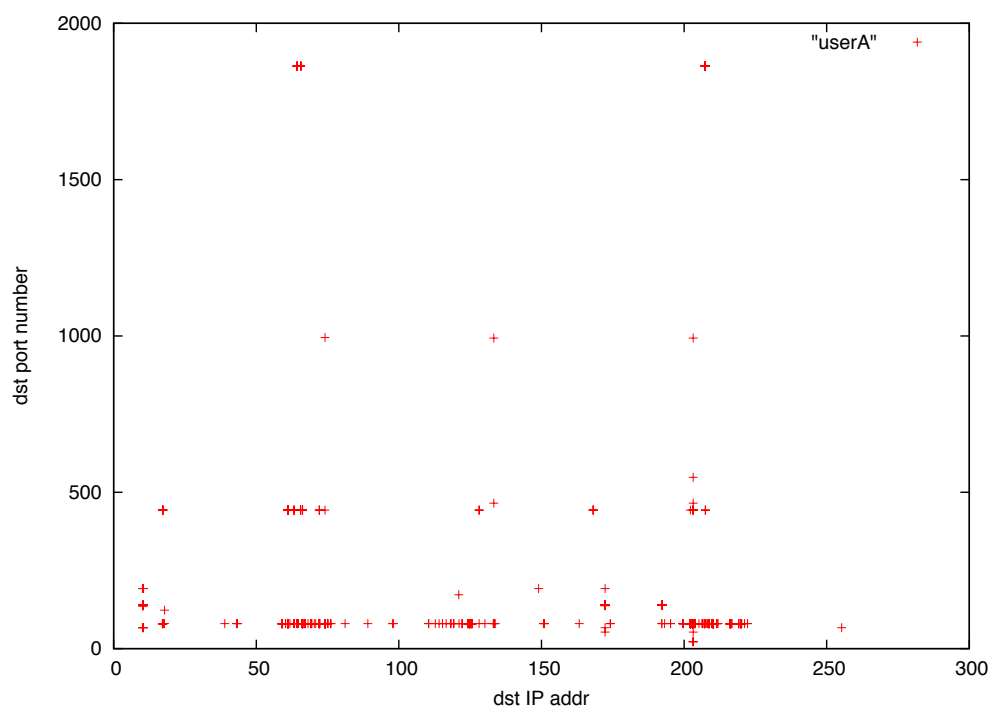


図 4.3: ホスト A の送信先 IP アドレス・ポート番号

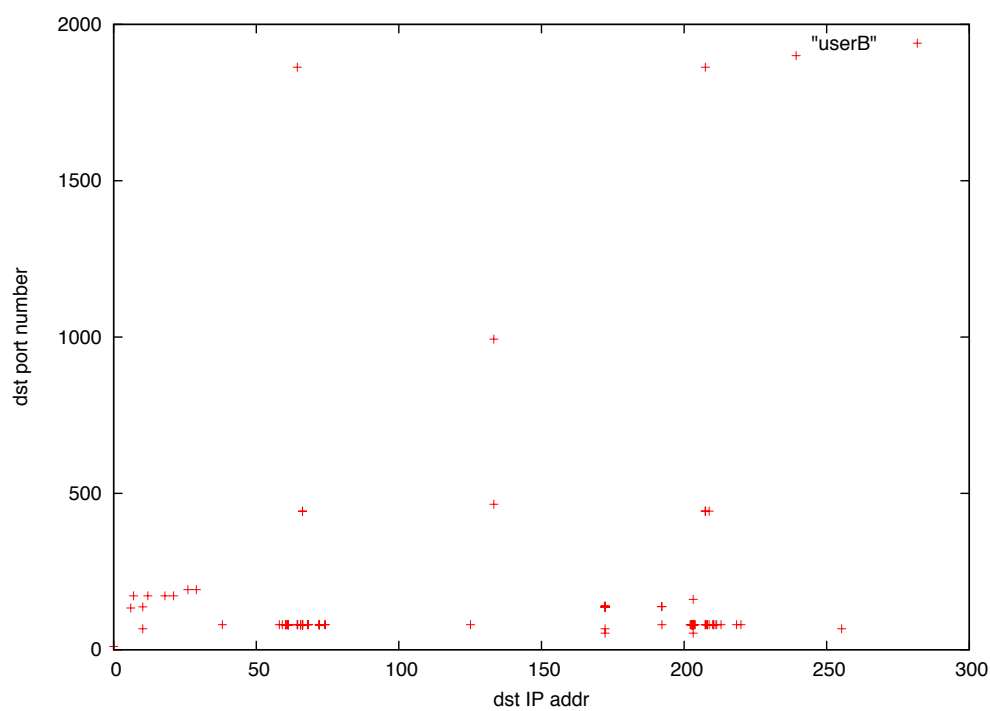


図 4.4: ユーザ B の送信先 IP アドレス・ポート番号

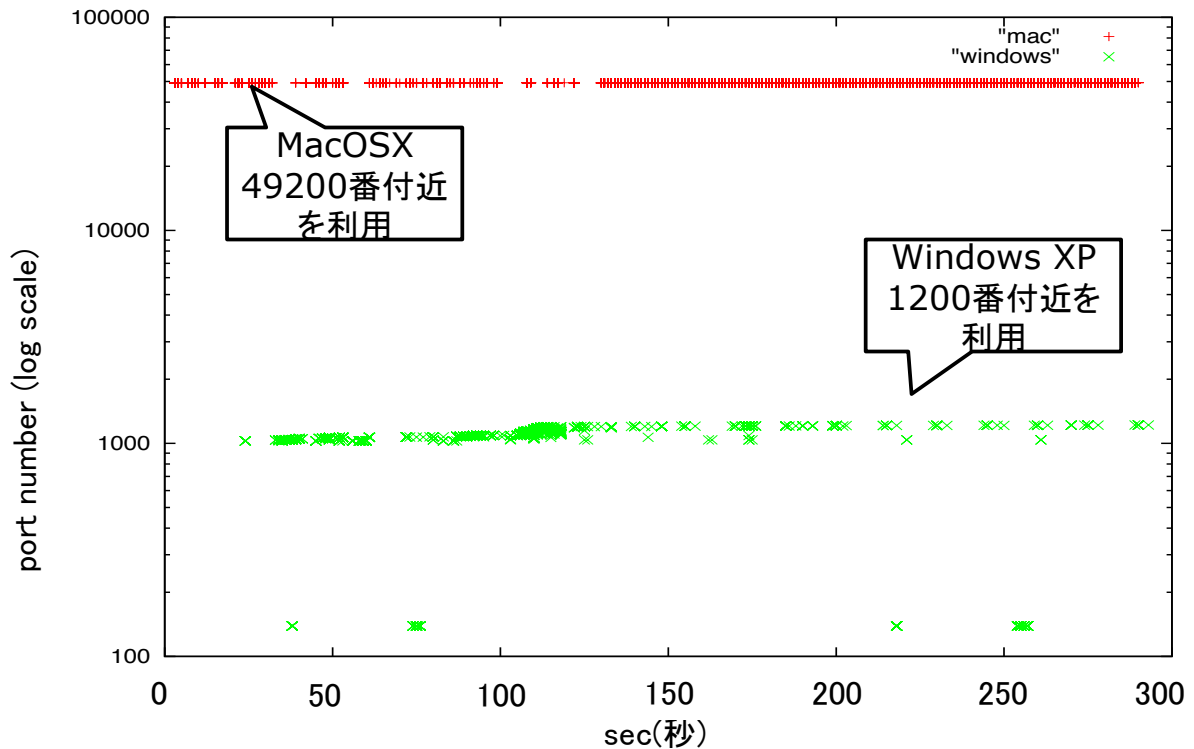


図 4.5: MacOSX と WindowsXP の利用送信元ポート番号

● 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

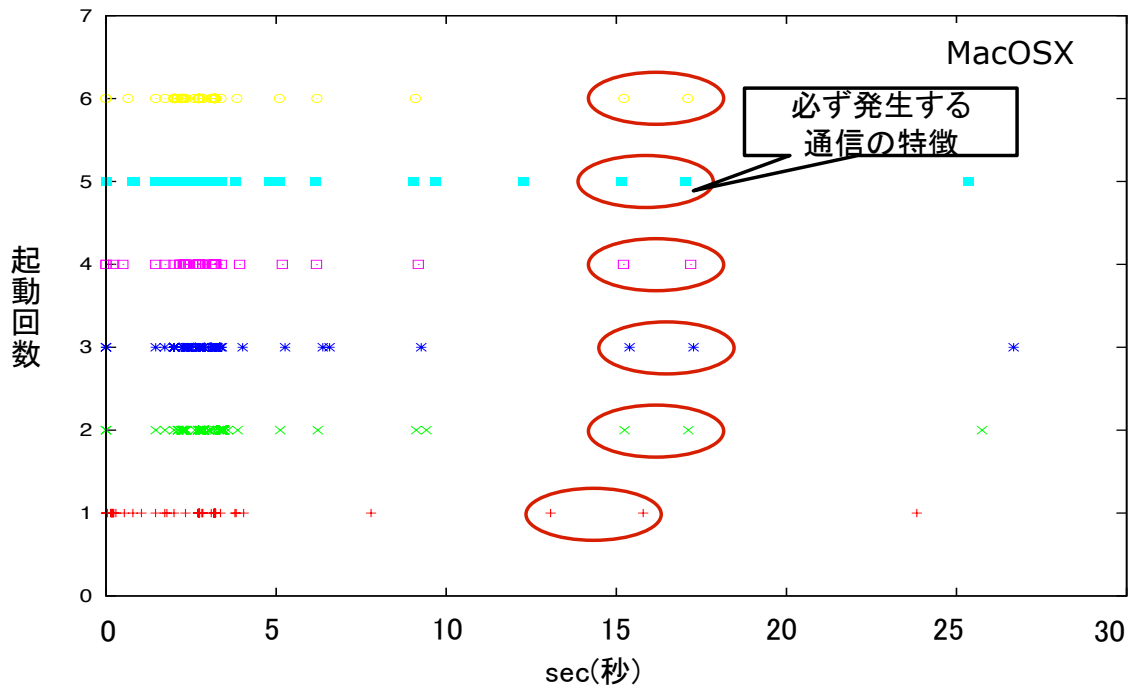


図 4.6: MacOSX の起動時のパケットの発信タイミング

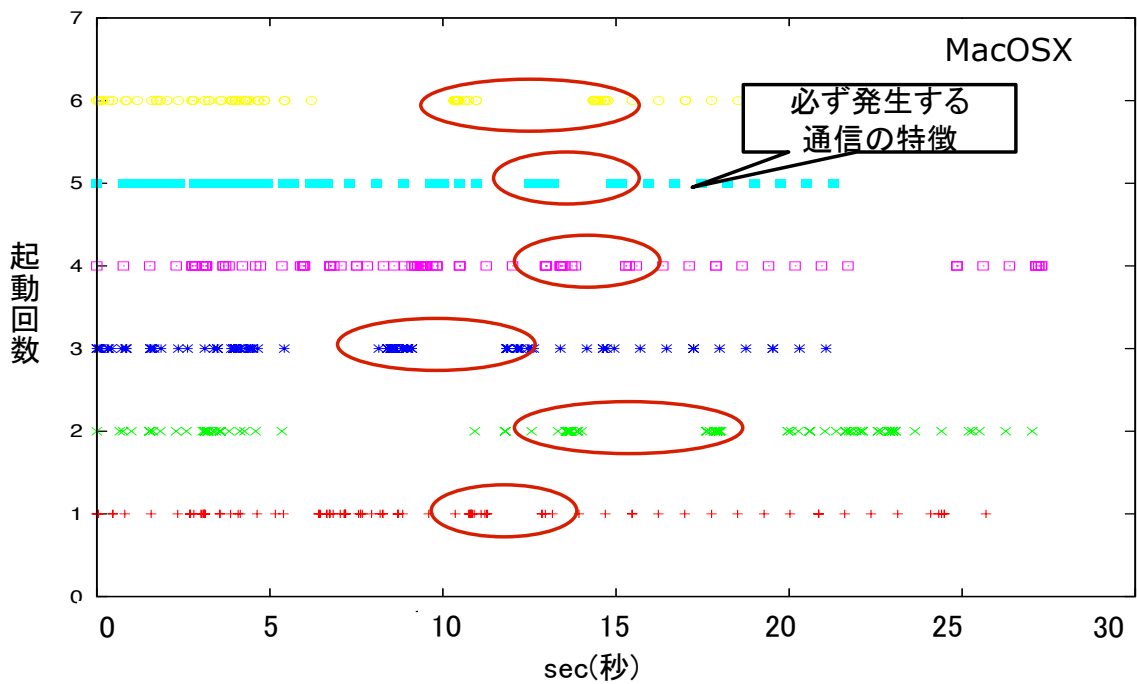


図 4.7: Windows Vista の起動時のパケットの発信タイミング

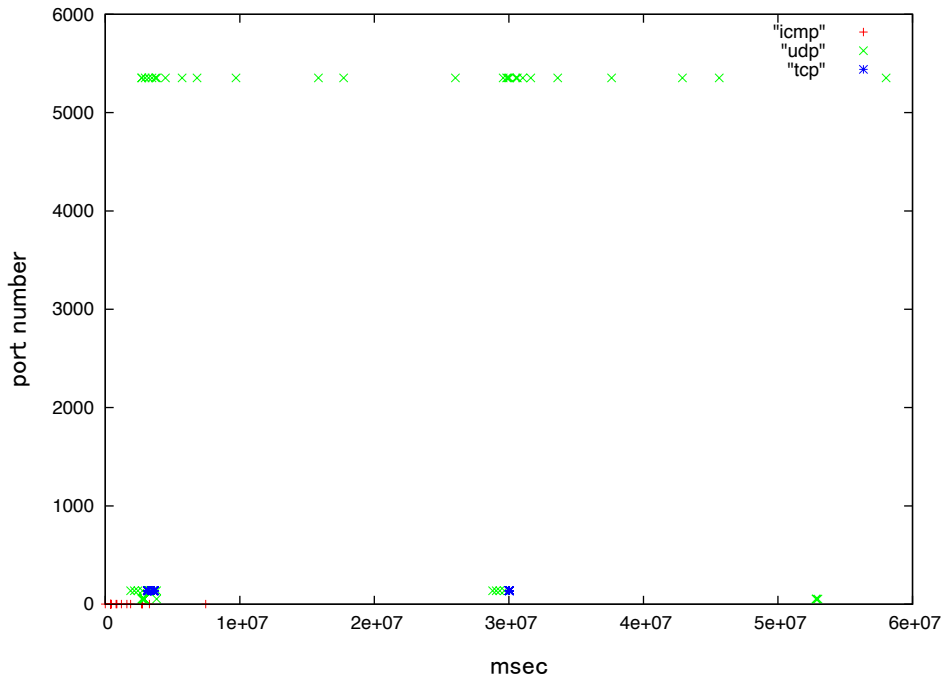


図 4.8: MacOSX のポート番号とプロトコル別パケットの発信タイミング

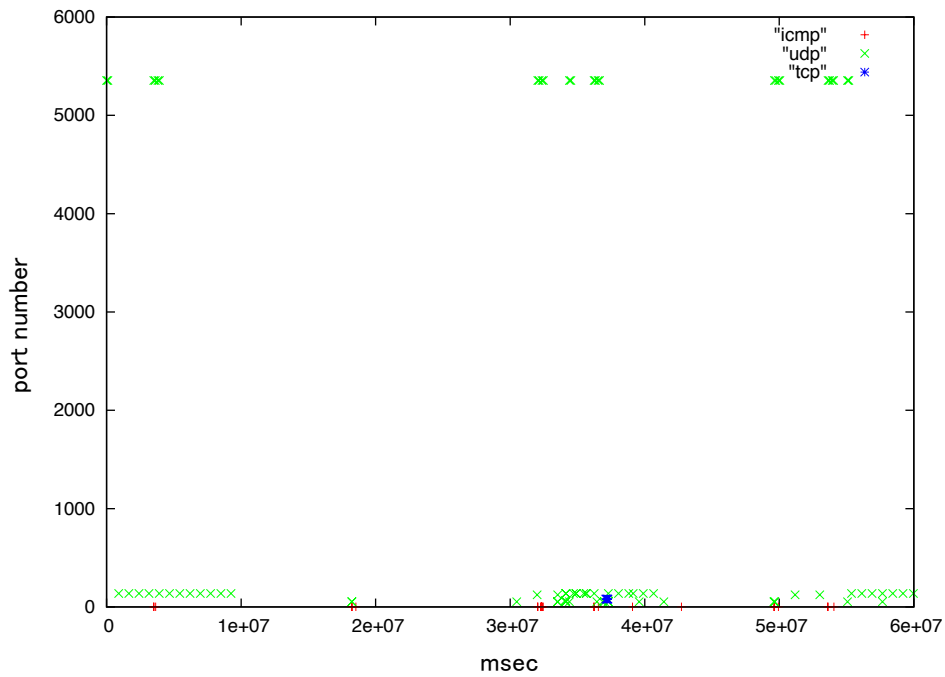


図 4.9: Windows Vista のポート番号とプロトコル別パケットの発信タイミング

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 メッセージャを対象として取り上げる。

IMAP や MSN メッセー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 メッセー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

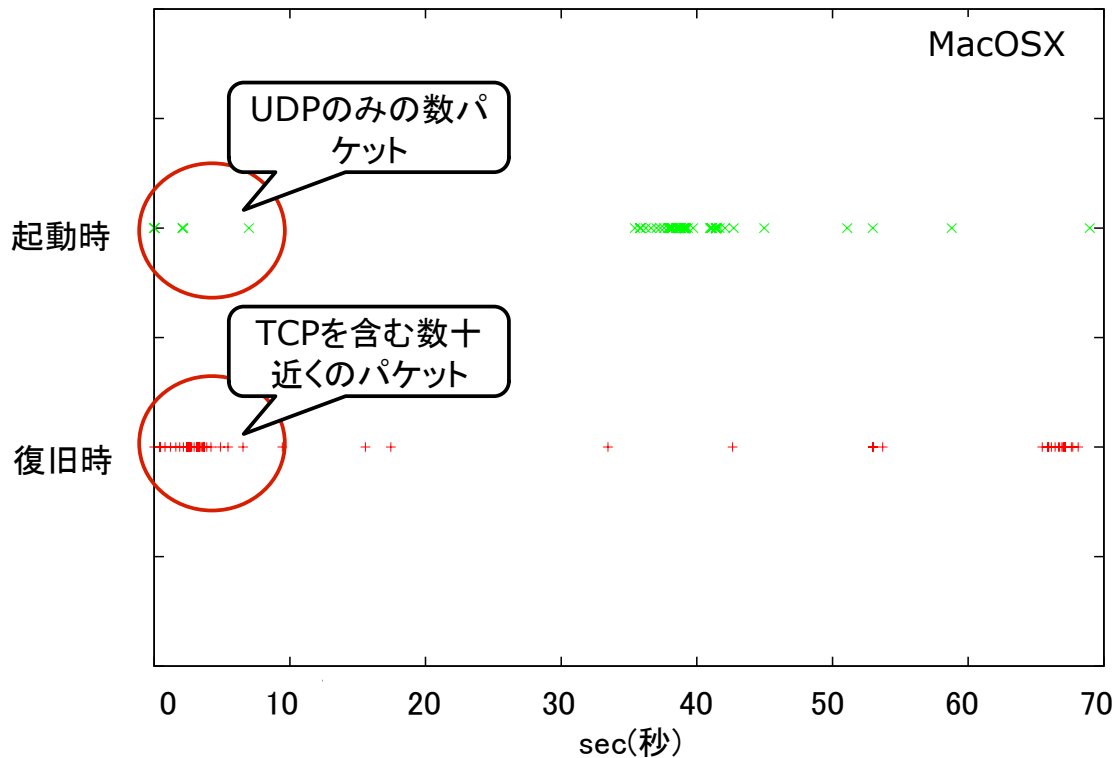


図 4.10: 起動時と復旧時におけるパケットの発信タイミングの比較

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があるため、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[36] や Twitter[37] 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
Omunigrafile	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	メディアプレイヤー
Quick Time	メディアプレイヤー
VLC media player	メディアプレイヤー
Windows media player	メディアプレイヤー

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정에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

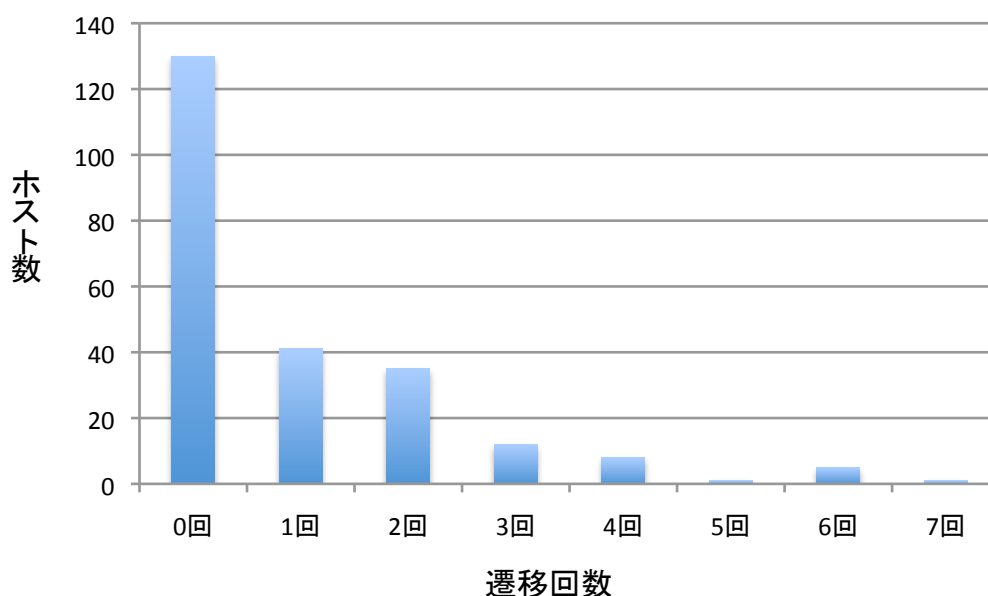


図 4.11: 時間におけるホストの IP アドレス遷移回数

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ー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ーが脅かされる点である。

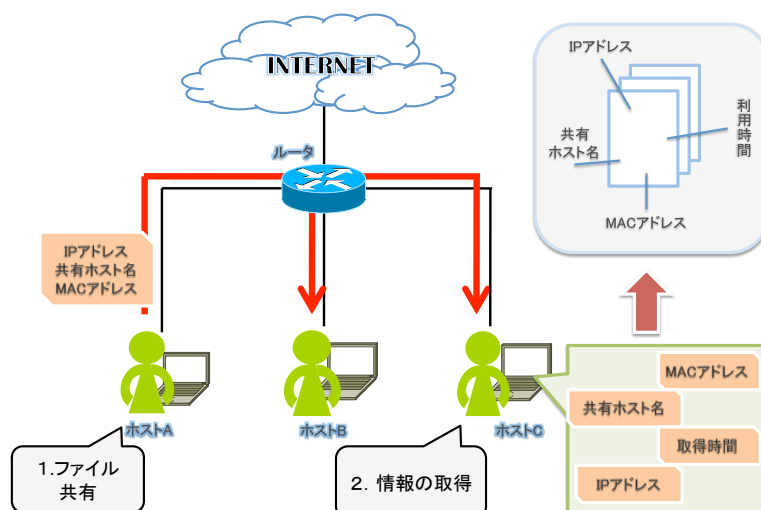


図 4.12: パケットヘッダ情報の収集

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ー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。

mDNS 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 という名前を設定されている。し

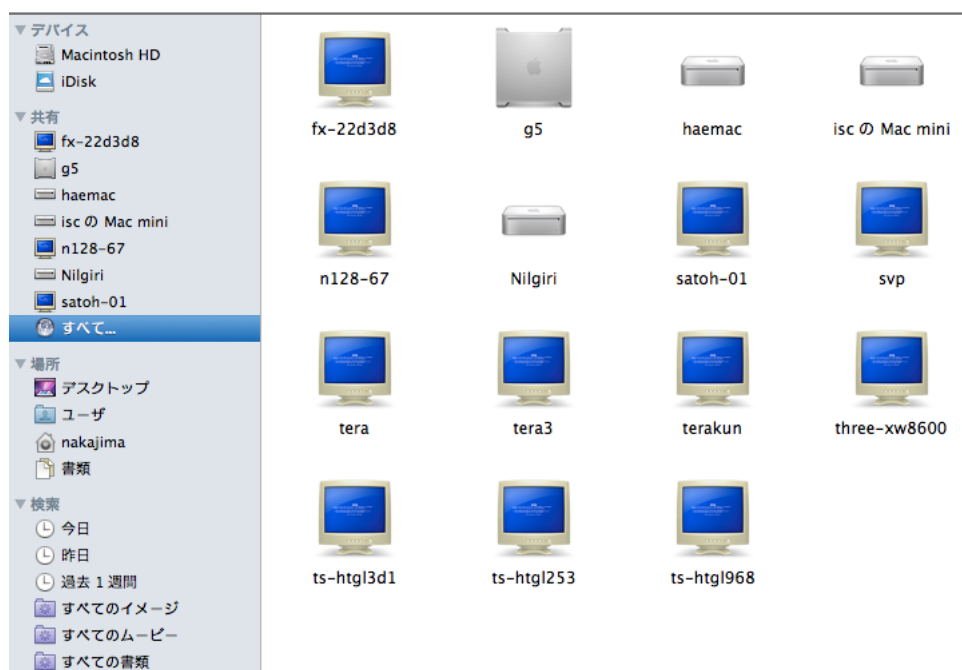


図 4.13: MacOSX の Finder

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ー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、プライバシー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 2.4.4 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 [38]。

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、携帯電話など小さな機器までも 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。

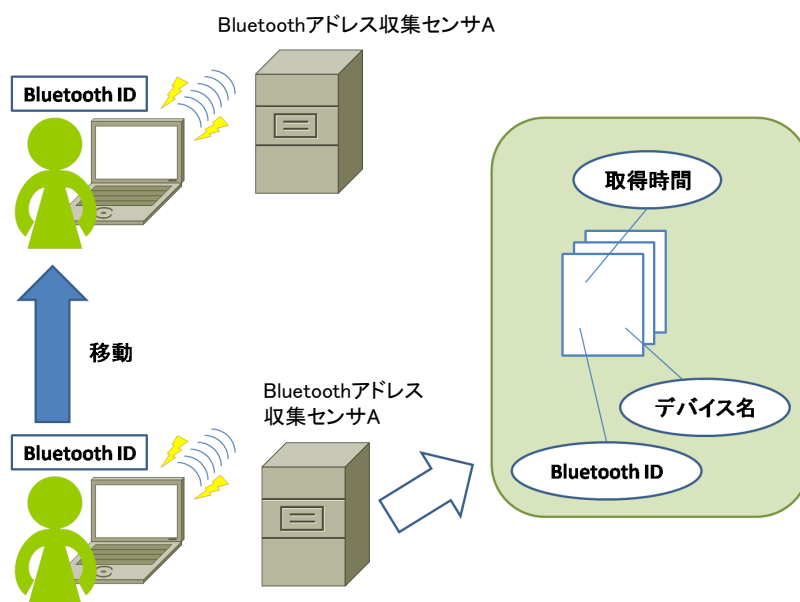


図 4.14: Bluetooth 情報の収集

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ー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

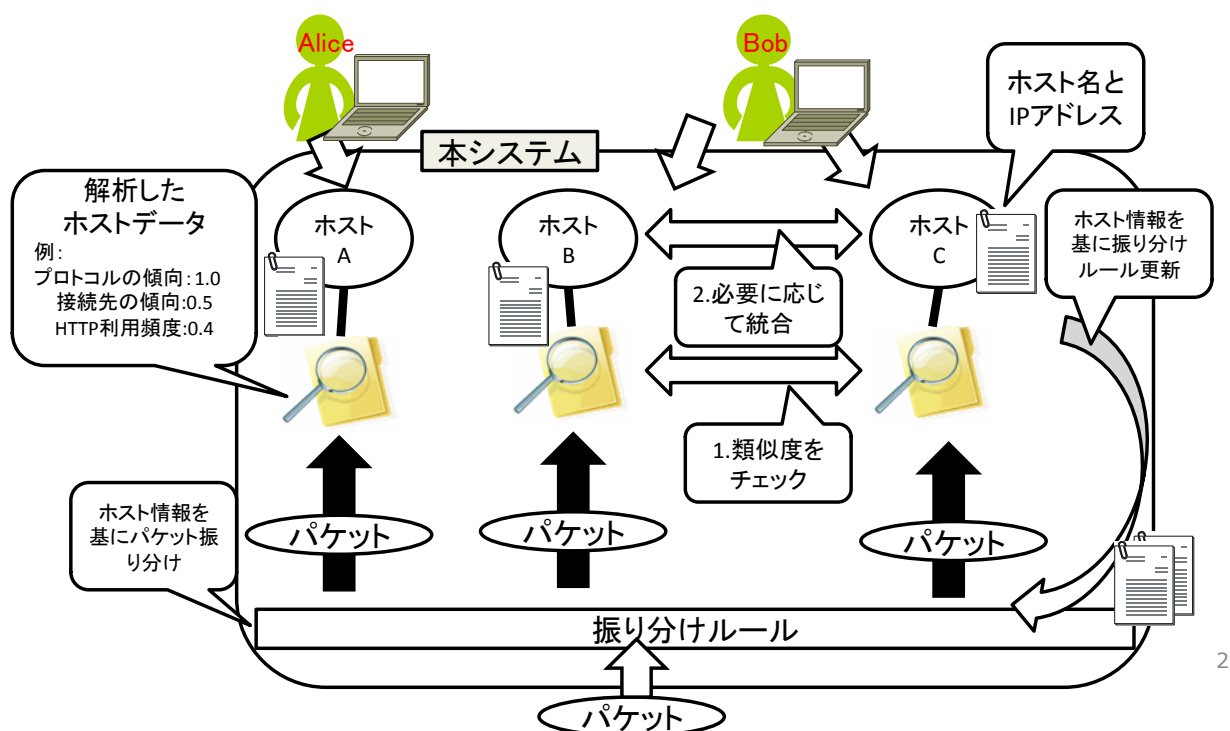


図 5.1: システムの動作概要

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

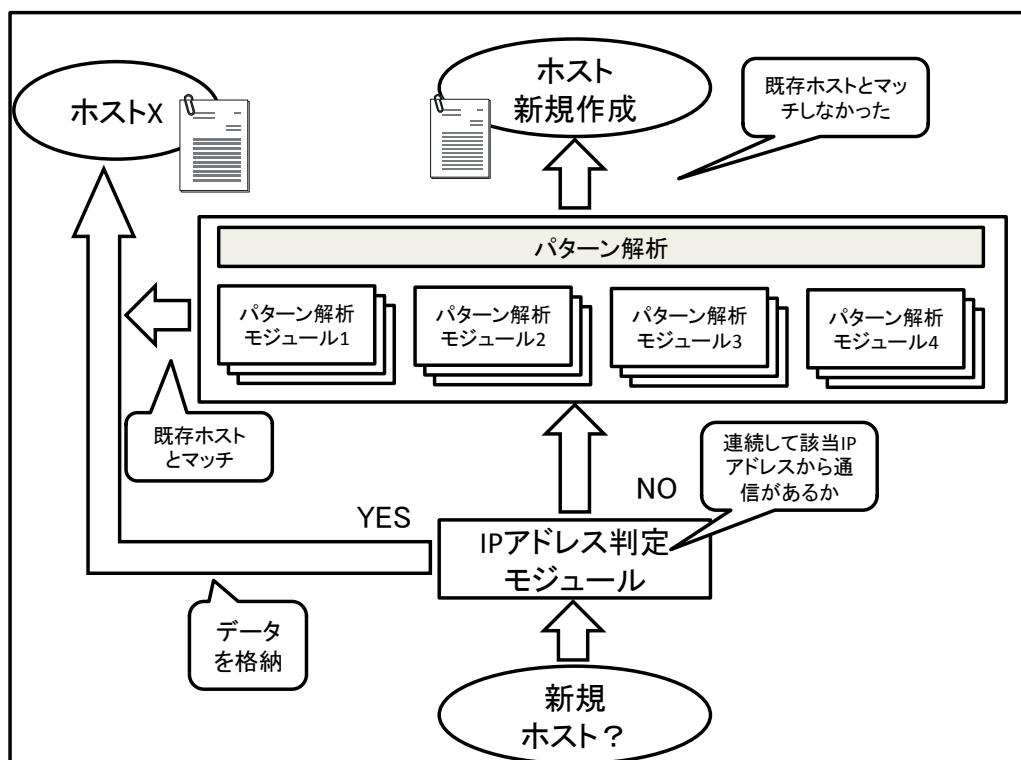


図 5.2: ホスト識別手法の設計

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。これら一覽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 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。

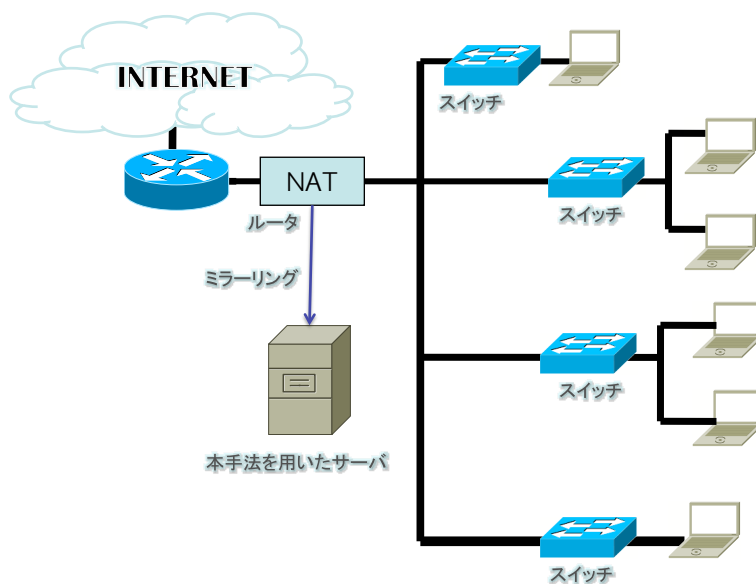


図 5.3: WIDE 合宿ネットワークトポロジ図

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 メッセージ	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、メッセージや IRC を利用しているユーザを分類した。メッセージを利用する場合、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。メッセージ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 メッセージ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[39]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しまいうため、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

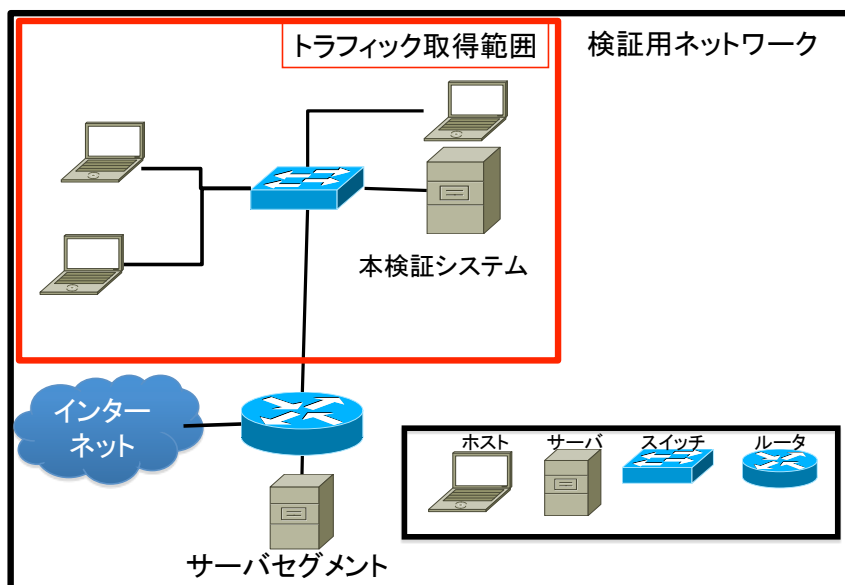


図 5.4: 共有ホスト名を用いた実験のネットワーク概要図

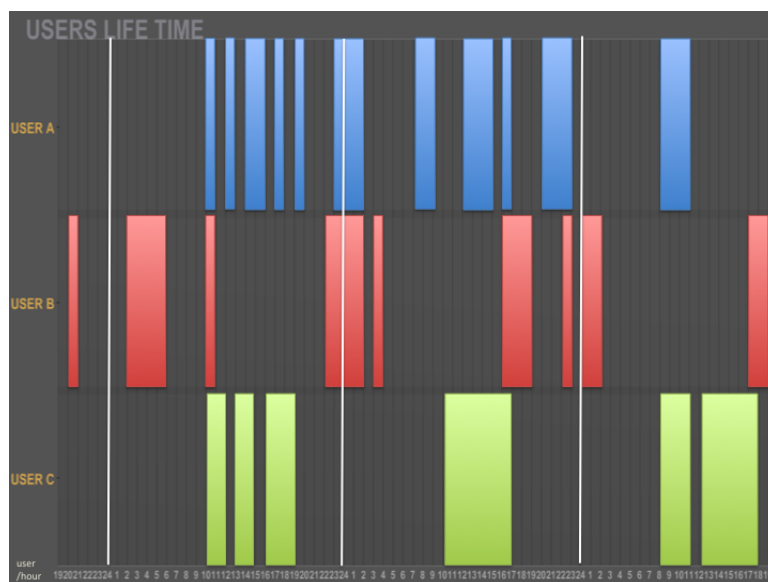


図 5.5: 共有ホスト名を用いたユーザ生活モデル

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ー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ことで、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[40] 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


```

<terminated> BtSearch [Java Application] /System/Library/Frameworks
BlueCove version 2.1.0 on mac
2009-12-15 22:23:20 001F5B72B520 haemac
2009-12-15 22:23:34 001FF3A87BCD amigasa
2009-12-15 22:23:39 002500C03716 16core
2009-12-15 22:23:41 002608D0E009 nyasnow
2009-12-15 22:23:44 00236CA360D7 Rize
2009-12-15 22:23:46 0026B0F8F99C isc の Mac mini
2009-12-15 22:23:52 00250085E57B tp1
2009-12-15 22:23:58 001F5BE37D6F Fran

```

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 2.4.4 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 は昼に研究室を訪れている。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ー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

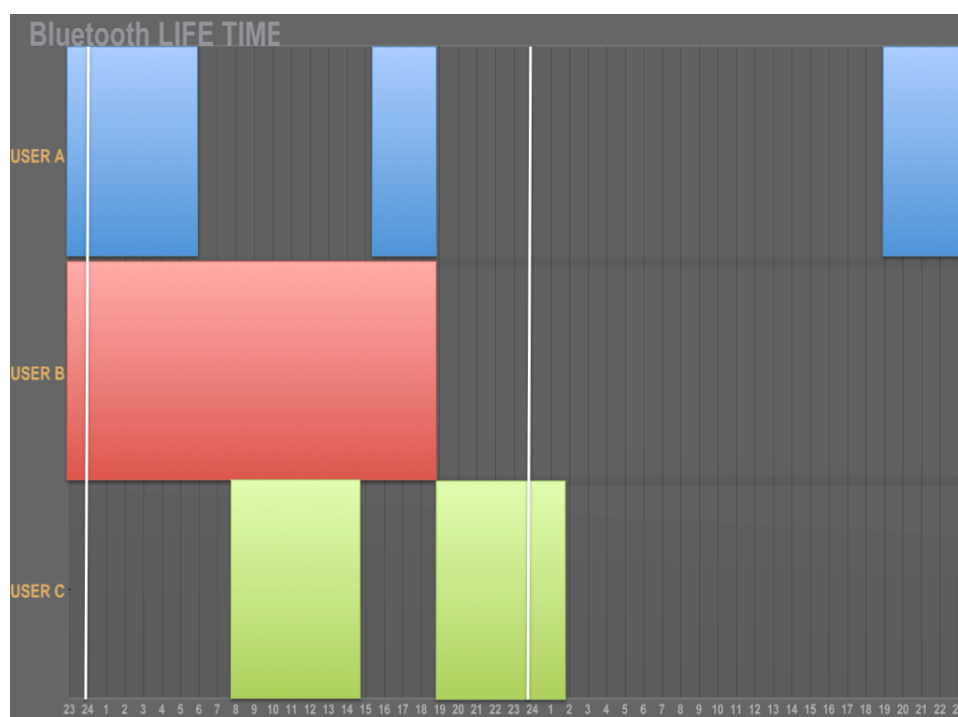


図 5.7: Bluetooth デバイスの検出によるユーザのライフタイム

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

プロフィールを作成する。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ィールが困難である。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、条件を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ィー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ー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ー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ー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ー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ー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、ユーザのプライバシー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イバシーを保護することができる。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ー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、プライバシー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ーを脅かすため、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、プライバシー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、容易に管理できることが必須である。また、ユーザ自身が発信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ー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ー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[41]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ー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、プライバシー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ーを保護するために、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ー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ー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、プライバシー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ー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、自身のプライバシー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ー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、プロフィー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ィールの作成ができた。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ー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ー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イバシー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ー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ー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、新しいプライバシー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ー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ー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ー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ー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ー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ーの問題に取り組み、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ー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ー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ー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イバシー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ーに関する調査や、複数の情報を組み合わせるアルゴリズムの提案をによって、プライバシー情報を判断するシステムを構築することを通して、今後のプライバシー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 齊藤 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 巖氏、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 -ファッション、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. 個人情報保護に関する法律. <http://law.e-gov.go.jp/htmldata/H15/H15H0057.html>, 12 2009.
- [4] Alan Westin. *Privacy and Freedom*. New York Atheneum, 1967.
- [5] Mozilla Japan. 次世代ブラウザ firefox. <http://mozilla.jp/firefox/>, 1 2010.
- [6] Adobe Systems Incorporated. Adobe. <http://www.adobe.com/>, 12 2009.
- [7] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris J. Hoofnagle. Flash cookies and privacy. *SSRN eLibrary*, 2009.
- [8] Quantcast Corporation. Quantcast - home. <http://www.quantcast.com/>, 12 2009.
- [9] 山崎賢児 and 勅使河原海. 高度なパーソナライズ実現のための統合サービスエージェントの設計. *IPSSJ SIG Technical Report*, pages 105–110, 3 2005.
- [10] 衣袋宏美. ネット視聴率白書 2008-2009. 翔泳社, 2008.
- [11] NebAD Inc. *NebAD Inc*. <http://www.nebuad.com>, 6 2008.
- [12] Annie De Montigny-Leboeuf. Passive network discovery for real time situation awareness. 4 2004.
- [13] T. KARAGIANNIS. Blinc : Multilevel traffic classification in the dark. *ACM Sigcomm, Philadelphia, PA, Aug. 2005*, 2005.
- [14] PASMO co. Pasma. <http://www.pasmo.co.jp/>, 12 2009.
- [15] East Japan Railway Company. Jr 東日本:suica. <http://www.pasmo.co.jp/>, 2009 10.
- [16] Google. Google. <http://www.google.com>, 12 2009.
- [17] Samuel D. vWarren and Louis D. Brandeis. The right to privacy. *Harvard Law Review*, 4(5):193–220, December 1890.

-
- [18] JIPDEC. プライバシマーク制度. <http://privacymark.jp/index.html>, 12 2009.
- [19] Pala Inc. ぷららホームページ. <http://www.plala.or.jp>, 11 2009.
- [20] NTTPlala. ニュースリリース (2006.6.13). http://www.plala.or.jp/member/option_service/secuplus/nbb/start_wfs.html, 12 2009.
- [21] 村本俊祐, 上土井陽子, and 若林真一. k-匿名性を利用したデータ一般化によるプライバシー保護. *DEWS2007*, 2007.
- [22] 総務省. **電波法 第 5 章**, 12 2009.
- [23] 財団法人インターネット協会. **インターネット白書 2008**. 株式会社インプレス R&D, 6 2008.
- [24] 小澤 勇騎, 大東 俊博, and 森井 昌克. 無線 lan 暗号化方式 wpa-tkip の脆弱性とそれを用いた攻撃方法の提案. *Computer Security Symposium 2009*, 2:805–810, 10 2009.
- [25] IVT Corporation. Bluesoleil, the most easy-to-use, innovative and interoperable bluetooth software, bluetooth driver. <http://www.bluesoleil.com/>, 12 2009.
- [26] 上原雄貴, 水谷正慶, 武田圭史, and 村井純. セキュリティインシデント対応のためのユーザ特定支援システムの実装. **情報処理学会 第 71 回全国大会**, 3 2009.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] Electronic Frontier Foundation. Panopticlick. <http://panopticlick.eff.org/>, 1 2010.
- [31] 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.
- [32] 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.
- [33] A. Tootoonchian, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr: Better privacy for social networks. *CoNEXT*, pages 169–180, 2009.

- [34] 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.
- [35] M. Zalewski and OS Passive. Fingerprinting tool. <http://lcamtuf.coredump.cx/p0f.shtml>, 1 2010.
- [36] Inc mixi. [mixi]. <http://mixi.jp>, 9 2009.
- [37] Twitter. twitter. <http://twitter.com>, 9 2009.
- [38] 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.
- [39] Evernote Corporation. Welcome to your notable world — evernote corporation. <http://www.evernote.com/>, 1 2010.
- [40] Apple. Apple - download music and more with itunes. play it all on ipod. <http://www.apple.com/itunes/>, 12 2009.
- [41] 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.