

ネットワーク通信情報の収集によるユーザ追跡

上原 雄貴^{1,a)} 水谷 正慶² 武田 圭史³ 村井 純³

受付日 2011年5月23日, 採録日 2011年11月7日

概要: 近年, インターネットを介した各種サービスにおいて, ユーザごとの行動履歴や趣味趣向に応じた最適化されたサービスを適時に提供することで, 新たな付加価値を生み出すための取組みが行われている. 実際にこのようなサービスを利用するためには利用者が専用のアカウントを登録するといった特別な操作が必要である. また, 従来ホストの識別に IP アドレスや MAC アドレスが利用されているが, IP アドレスはネットワーク環境によって変化し, MAC アドレスは詐称可能であるため, 継続的にホストや所有者を特定できない. そこで本研究では, 特定のユーザが利用するデバイスごとに, ユーザの利用形態によって特徴的なトラフィックパターンが現れるかを事前に調査した. そして, その結果を用いてネットワーク上に発信されているパケットのヘッダ情報を収集, 解析することでホストおよびユーザを特定する手法を提案した. 本提案手法において, ユーザが発信するパケットのペイロード情報はプライバシー侵害の懸念があるため, パケットヘッダ情報のみを収集することでホストおよびユーザを特定することを目的とする. そして, 200 人ほどが参加するカンファレンスや学術ネットワークにおいてユーザの許諾を得たうえで実験を行い, ホストやユーザを本手法によって特定可能であることを示した. これによって, サービス提供者はユーザに特別な操作や設定などを求めることやデバイスの制約などを受けることなく, ユーザごとに特化したサービスを提供できる可能性を示した.

キーワード: ネットワーク監視, フォレンジック, プライバシ, インターネット

User Tracking by Network Communication Information Surveillance

YUKI UEHARA^{1,a)} MASAYOSHI MIZUTANI² KEIJI TAKEDA³ JUN MUARAI³

Received: May 23, 2011, Accepted: November 7, 2011

Abstract: As demand for internet services specifically customized for each user increases, service providers make effort to build extra value on their services by providing useful services, based on personal action history and personal interest. However, in order to use these services, they require users to install specific application. IP addresses and MAC addresses are being used for host identification, but IP addresses vary as the hosts' network environment change, and MAC addresses are deceivable. Therefore, continually specifying hosts and users is impossible. In this paper, we propose a method to identify and track users by collecting and analyzing network traffic, which enables users to receive customized services without installing special application nor user of specific devices. Payloads of packets, transmitted by users, are subjected to privacy, so in the method, host and user identification is performed by collecting information from packet header. This methods' practicability has been verified through experiments, performed on a conference and an academic network. Through the experiment, possibility for service providers to offer users user specific services, without requiring users to perform particular actions nor setting, nor limiting users to use specific devices, has been shown.

Keywords: network monitoring, forensic, privacy, internet

1. はじめに

近年、情報技術の発展によってユーザ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、解析することにより、ユーザの要求に応じたサービスの提供が実現した。特にユーザの振舞いに応じたデータを収集することで、レコメンドサービスなど様々なサービスや取組みが行われている。しかし、そのようなサービスはユーザ自身にアカウント登録などの特別な操作が必要となる。また、ホストを識別する際には、従来 IP アドレスや MAC アドレス、Cookie などを用いるが、IP アドレスは可変であり、MAC アドレスや Cookie などの識別要素は容易に変更、詐称可能である。そこで、本研究は特定のユーザが利用するデバイスはユーザの利用形態によって特徴的なトラフィックパターンが現れることに着目し、パケットのヘッダ情報を利用してホストとユーザを識別する手法を提案する。

本提案手法ではネットワーク管理者などネットワーク上のトラフィックを取得できるユーザが、法的観点やネットワークの制約上ペイロードを見ることができない状況での情報収集を想定する。本提案手法によって、個人情報や直接取り扱うことなくまた通信内容を直接的に取得せずにユーザが使用する機器やユーザを特定することが可能となる。

2. 関連研究

ユーザが発信する情報を利用して、ホストやユーザに関する情報を取得する様々な研究が存在する。ここではユーザが発信する情報をもとにホストやホスト利用者を特定する研究をあげる。

2.1 ネットワーク管理者がユーザが発信する情報を収集・解析する手法

受動的にトラフィックを収集、解析する研究において、ネットワークトラフィックに着目することでネットワークの状態やプロトコルを把握する研究に Karagiannis らの研究 [1] や、Montigny-Leboeuf の研究 [2] があげられる。また、和泉らはパケットのヘッダ情報を利用してネットワーク上に存在するホストが利用するアプリケーションを識別する手法 [3] を提案している。しかし、これらの研究の目

的はネットワーク状態の把握であり、ホストやユーザを特定する本論文の目的とは異なる。

2.2 Web 上の情報とホストを結びつけるユーザ特定手法

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を調査した研究に Krishnamurthy らによる研究 [4] があげられる。この研究はユーザが SNS に登録する情報と SNS から発行される Third-party Cookie を解析することで、個人と IP アドレスを結びつける手法とそのリスクを提唱している。他にも、松尾らによる Web 検索エンジンを利用したユーザプロファイル手法 [5] や、本村らによるネットワーク上の情報を利用して、プライバシーを侵害や対策についた手法 [6] が提案されている。しかし、これらの研究は対象ユーザが Web 上に情報を公開していない場合はこの手法は利用できず、汎用性に欠ける。

2.3 ブラウザ情報を利用したホスト識別

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検証する研究に Electronic Frontier Foundation による研究 [7] があげられる。この研究では HTTP における 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、ホストを識別する。そして、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を公開することで、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。ブラウザ情報といった普段意識されない情報を利用してホストを識別する方法は、本研究と類似するものがあるが、特定のブラウザしか情報を取得できないことや、HTTP を利用しないホストの識別ができないという制約がある。

3. デジタル情報の収集によるホスト、ホスト利用者の特定とその課題

デジタル情報を発信する媒体は、パソコン、携帯電話など多岐にわたる。特にインターネットの発展はすさまじく、日本におけるインターネット人口は 1997 年の 572 万人から、10 年後の 2007 年には 8,227 万人 [8] と爆発的に増加し、さらに増え続けている。このことから、多くのユーザが端末を通じて情報を頻繁に発信しているといえる。そして、ユーザがデバイスを持ち歩くようになり、ユーザに特化した情報を提供するサービスの需要が高まっている。たとえば、情報のレコメンデーションやユーザのライフログなどがあげられる。しかし、これらのサービスを利用するためにはユーザアカウント登録や、専用デバイスの利用などの制約が存在する。そのため、サービス提供者はユーザに特別な操作や設定などを求めることなくサービスを提供することが求められている。従来、機器の識別には IP アドレス、MAC アドレスや Cookie などの識別子が利用されて

¹ 慶應義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University,
Fujisawa, Kanagawa 252-0882, Japan

² 日本アイ・ビー・エム株式会社東京基礎研究所
IBM Research, Tokyo Research Laboratory, Yamato,
Kanagawa 242-8502, Japan

³ 慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252-0882, Japan

a) nakajima@sfc.wide.ad.jp

いる。しかし、IP アドレスは可変であり、Cookie などを利用したホストの識別においても、ホストが Cookie を受け入れない場合や、ユーザが Cookie を削除することができるため、継続的にホストを追跡することが難しい。また、機器の識別子として利用されている MAC アドレスは詐称が容易に可能である。さらに、近年普及しているノート PC などは有線や無線といった 1 つの機器に複数の MAC アドレスが付与されている場合が多い。そのため、有線と無線のインタフェースを持つ 1 台の機器を複数の機器として判断してしまう問題がある。これらの問題に対応するためには、IP アドレスや MAC アドレス、Cookie などの識別子以外で機器を特定できる方法が必要である。

本論文ではユーザが普段インターネット上に発信しているパケット情報を、ネットワーク管理者が観測し、組み合わせることでホストやユーザを特定する手法を提案する。本提案手法で利用する情報は、普段ホストが送信しているパケットヘッダ情報である。2010 年度総務省において Deep Packet Inspection によるターゲティング広告の技術が取り上げられた際 [9] に、ペイロードを取得することについて議論が巻き起こっている。そのため、ペイロードは取得せずにパケットヘッダ情報のみ取得することで、この制約を外すことや、ペイロードの取得に比べてユーザの同意が得ることが容易となる。本論文で述べるパケットヘッダ情報とは、IP、TCP、UDP、ICMP ヘッダのことを指す。そして、ネットワーク管理者は、これらの情報を対象の組織内 LAN 内などの管理ネットワークで取得し、本手法を用いることを想定している。これによってユーザは意識することなくサービスを受けることができ、サービス提供側やネットワーク管理者はユーザに操作を要求することなく、より正確に管理ネットワークに参加している機器やユーザ情報を得ることができる。

また、該当するホスト所有者がネットワークを利用しているという席情報を用いて、ホストの所有者を特定する手法も提案する。この手法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するユーザでも実現可能である。しかし、複数のセグメントが存在した場合など、ネットワークの構成によって利用できない場合があるため、本論文ではこの 2 つの手法を組み合わせることでホストとユーザを特定する。図 1 に本提案手法の概要を示す。まず、ネットワークトラフィックから、パケットヘッダ情報のみを用いて各ホストに関する情報を収集し、ホストごとの振舞いを記録したプロファイルを作成する。次に、プロファイルをもとに特定対象ホストを特定する。そして、特定対象ホストの所有者がネットワークを利用している情報を収集し、パケットヘッダ情報により取得したホストプロファイルと組み合わせることでホスト利用者を特定する。

これら 2 つの手法によって、ネットワーク管理者はネットワークへの不正ホストの接続の監視、追跡可能性の検証

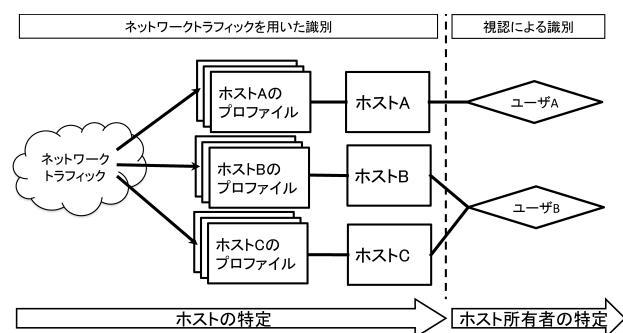


図 1 本提案手法の概要図

Fig. 1 Schematic diagram of proposed technique.

や、ユーザごとに適したサービスやオプトインによる広告の利用が容易になる。

4. パケットヘッダ情報によるプロファイル手法

ネットワーク管理者は、管理しているネットワークにおけるユーザの発信するパケットの情報を容易に取得できる。一般的なネットワークでは、セキュリティ対策やネットワークの運用の関係上、管理ネットワークと外部のネットワークとの中継点でトラフィックを監視できる場合が多い。しかし、ペイロードを取得する場合は電気通信事業者法などの観点からユーザ全員の同意を得なければならない。そこで、ペイロードを除いたパケットのヘッダ情報からホストを特定する手法を提案する。

4.1 事前実験

特定のユーザが情報を利用するデバイスにはユーザの利用形態によって発信する情報に特徴的なトラフィックパターンが現れる可能性が高い。そのトラフィックパターンを取得、解析することによって、デバイスの追跡が可能かどうかの事前実験を行った。ユーザが情報を利用するデバイスの特徴は、対象のネットワークによって異なることが推測できる。そのため、今回の事前実験においては、200 人程度が在籍する筆者の所属する研究室の学術ネットワークを対象とする。このネットワークは外部と通信を行う際に、必ず本手法を適用した機器を通過する構成であり、IP アドレスは DHCP によって割り振られた IP アドレスと固定 IP アドレスが混在し、ローカル IP アドレスの重複や NAT 構成がされていないネットワークである。その環境下で、無作為に選出したホスト 3 台の通信を観測した。そして、実験で取得した各項目を利用することによって、どの程度ホストを絞り込むことができるかを検証した。結果を表 1 に示す。

本ネットワークにおける事前実験において、発生頻度が高く取得が容易な通信と、発生頻度が低いが取得すると絞り込みが容易になる通信の 2 種類に分類できることが分か

表 1 事前に作成したホストのプロファイルデータ

Table 1 Host profile data produced in advance.

項目	ホスト A 候補数	ホスト B 候補数	ホスト C 候補
接続したメールサーバ群	62	2	32
接続した SSH サーバ	13	0	16
OS	20	22	18
特徴的な挙動	0	23	0
接続した IRC	0	0	0
起動時のアプリケーション挙動	18	0	18
アプリケーションアップデートの通信	2	0	0

る。まず、発生頻度の高い識別要素にメールサーバの接続先情報があげられる。メールサービスは多くのユーザが利用し、複数メールサーバに接続しているホストは容易に絞り込みが可能であることが分かった。ホスト候補 B, C は複数のメールサーバにアクセスしており、メールサービスの利用および接続先サーバを識別要素とすることで、ホスト B は 2 台まで、ホスト C は 32 台まで絞り込むことができた。そのため、取得が容易かつ判別能力は高いといえる。次に、リモートログインもコミュニティによっては比較的使われやすいサービスである。取得の容易さはメールサーバ接続情報と比較した場合には劣るが、容易に取得でき、ホストごとに差異が出やすいことが分かった。ホスト A, C は SSH を利用しており、メールと同じく接続先サーバを識別要素とすることで、ホスト A は 13 台、ホスト C は 16 台まで絞ることができた。このことから、メールよりも発生頻度は低いが、ホストの分類は容易にできるため、識別要素として利用できることが判明した。そして、OS 情報は Windows, MacOSX, Ubuntu であれば取得可能であるが、OS 情報はホストの絞り込みを十分にできなかった。そのため、メールやリモートログインと組み合わせることでホスト判別基準の 1 つとして利用することが望ましいことが分かった。一方で、発生頻度が低い、取得できると判別が容易な情報にホストのソフトウェアアップデートや特徴的な挙動があげられる。これらは取得することができれば絞り込みを容易にするが、アップデートであればポーリング、特徴的な挙動であればユーザがアプリケーションを動かす必要があるなど、発生頻度が低い。そのため、各項目における優先順位は低く設定する。同様に、現在利用人口が減っている IRC も優先順位を低く設定する。しかし、予備実験において、3 台の対象ホストは IRC を利用していなかったが、ネットワークにおいて IRC を利用するホストは 6 台存在したため、もしホストが IRC 利用している場合は強力な識別要素となりうることが判明した。また、起動時のアプリケーション挙動はアップデートと比べて取得は容易であるが、対象ネットワークに電源を切った状態で起動をする必要があるため、発生頻度は高いとはいえない。

事前実験において、前述した識別要素を利用し、学術

ネットワークでホストを識別した場合、200 人が参加するネットワークから、各 3 台のホストを 1 台に絞り込むことができた。このことから、事前実験を行ったネットワークにおいて、特定のユーザが情報を利用するデバイスは、ユーザの利用形態の特徴によって発信する情報に特徴的なトラフィックパターンが現れる可能性が高いことが分かった。同時に、事前実験で利用した情報の組み合わせる順番を、取得容易かつホストを判別できる要素が高い順に並べると、1. メールサーバ接続先情報、2. SSH サーバ接続先情報、3. OS 情報、4. 起動時のアプリケーション挙動、接続した IRC サーバ、アップデート情報、特徴的な挙動であった。これらの順番でホストを絞り込むことでホストの特定が可能である。また、ネットワークによって特定のユーザが情報を利用するデバイスは、ユーザの利用形態によって発信する特徴的なトラフィックパターンが変化する可能性が高いため、ISP などのユーザが異なるネットワークで利用する場合は、識別要素とする情報を再度検討する必要がある。

4.2 判別手法

4.1 節において、各ホストは利用状況に応じて、それぞれ特徴的なパケットを送受信していることが分かった。そこで、パケットのヘッダ情報からは、送信先、発信元 IP アドレス、ポート番号、発信タイミング、利用サービスなどの情報を取得、組み合わせることで、プロファイルを作成しホストを識別する手法を提案する。また、識別要素として取得する情報は、ネットワークやホストによって異なるが、今回は事前検証であげた大学や企業に所属するユーザが多く参加するネットワークでの識別要素について述べる。

図 2 にパケットヘッダ情報からホスト特定手法の概要を示す。ネットワークトラフィックから、各ホストの振舞いに関するデータを収集、蓄積する。本論文では、ホストの振舞いにより蓄積された情報をプロファイルと定義する。プロファイルとして蓄積される情報は、事前検証の情報と、ホストを継続して追跡するためにネットワーク接続時の状態情報を加えた以下の 5 つで識別する。

● OS の種類：発信元ポート番号

発信元ポート番号から、OS の推測や 1 度ネットワー

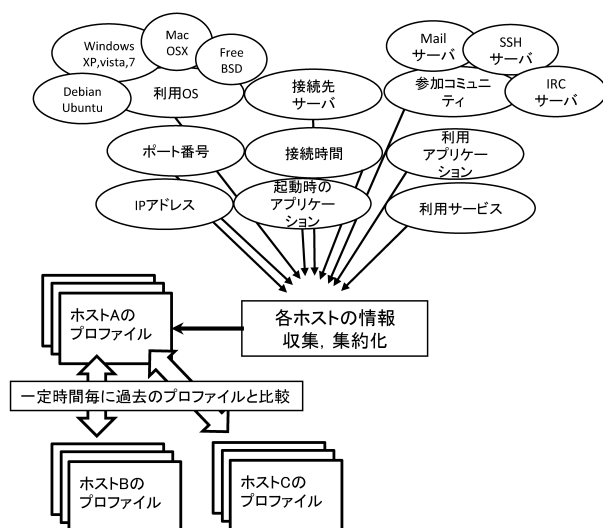


図2 パケットヘッダ情報によるプロフィール手法

Fig. 2 Profile method using packet header information.

クを離脱したホストを検索することができ、これらはユーザごとに異なるという特徴があるため識別要素として取得する。サスペンドから復旧したホストは、サスペンド前に利用されていた発信元ポートの数個前のポート番号を利用する傾向があり、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、再度追跡が可能である。また、発信元ポート番号からOS情報を取得することもできる。利用する発信元ポート番号はOSごとに異なり、Windows XPやUbuntuは特徴があり容易に識別できる。この技術はPassive Finger Printing [10]でも利用されている。また、OSアップデートの際の通信を取得することによって、MacOSXやCentOSなどの他OSを判別することも可能である。

● 利用サービス、アプリケーション

ユーザが利用するサービスやプロトコルも重要な識別要素となる。前提として、サービスを提供するサーバのアドレスブロックを取得していると、ホストが通信したサーバのIPアドレスからホストの利用しているサービスを特定する。そして、特定のプロトコル利用時の送信先IPアドレスとポート番号の組合せも利用する。今回のネットワークにおいて、識別要素とするプロトコルはSSH, IMAP, POP, VPN, IRCとする。他にも定期的な通信や、普段一般ユーザにとって利用されない通信ポートも十分な識別要素になる。

● ホストの所属するコミュニティ

利用サービス、アプリケーション利用時の接続先サーバも識別要素として利用できる。企業や大学に所属するユーザは、その組織が提供するサーバを利用するため、ユーザが利用したリモートログインサーバや

メールサーバを基準とすることで、ホスト利用者の所属する組織、団体ごとにホスト群が作成できる。本論文では接続先サーバごとに分類したホスト群をホストの所属するコミュニティと定義する。それらの情報を取得し、ホストのコミュニティを識別要素とすることでホストを判別する。また、接続先IPアドレスから位置情報や所属組織を類推するサービス [11] などと併用することでより正確にホスト利用者の所属組織を分類できる。

● 自動起動に登録されているアプリケーション

ホスト起動時にアプリケーションが立ち上がる動作もホストの識別基準となる。自動起動に設定されている場合、OSの起動とともにアプリケーションが特定の通信を行うため判断することができる。実際にホストが起動する際に発信するタイミングを計測したところ、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。その際の接続先サーバから自動起動に設定されているアプリケーションを推測し、ホストの識別要素とする。

● ネットワーク接続時の状態情報

ネットワークに接続したホストが電源を切った状態からの起動か、サスペンドからの復旧かを判定することはホストの特定において重要である。前述した自動起動に登録されているアプリケーションや、発信元ポート番号の情報を識別要素として利用する場合、ホストが起動か、サスペンドからの復旧かによって、利用情報が変化するためである。起動か復旧かの判別に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手法があげられる。図3はクリーンインストールしたMacOSXの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軸はIPアドレスが付与されてからの経過時間であり、Y軸は起動時および復旧時のホストの挙動のラベルである。この図から復旧時は、電源が落ちている状態からの起動時に比べて、IPアドレスが付与されてからすぐにTCPをはじめとしたパケットを送信している。つまり、サスペンド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ているため、起動して数秒で異なる挙動をした場合、サスペンドからの復旧と判断することができる。また、Windows, Ubuntu, CentOSも同様に検証したところ、接続時の復旧情報はこれら3つのOSにも利用可能であることが判明した。

これら5つの情報をもとに、ホストごとにプロフィールを作成する。ホストの特定は蓄積されたプロフィールをも

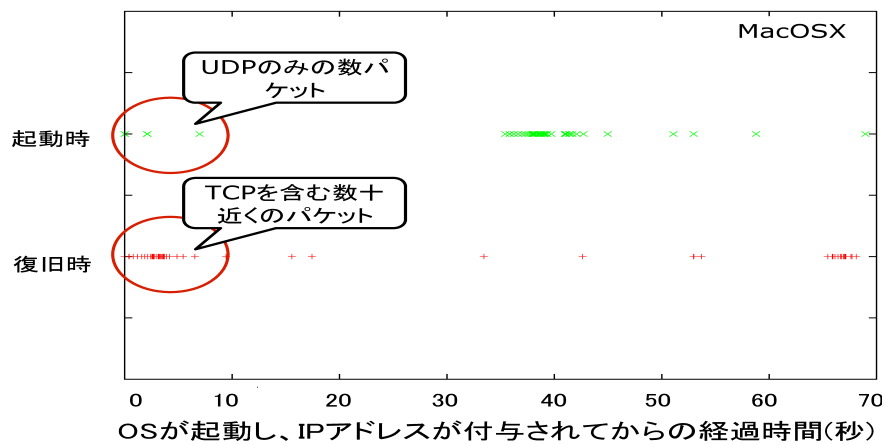


図 3 起動時と復旧時におけるパケットの発信タイミングの比較

Fig. 3 Comparison of submission time at wake-up time and restoration time.

表 2 本手法の検証環境

Table 2 Verification environment.

検証環境\要素	検証開始時刻	検証期間	ユーザ数	観測ホスト数	トラフィック量
カンファレンス	2009/9/07/17:44	約 47 時間	約 200 人	246 台	約 160 GB
研究室	2010/7/14/18:33	約 26 時間			
ネットワーク	2010/7/20/03:08	約 24 時間	約 200 人	192 台	約 62 GB

とに行うため、すべてのホストは最初の 30 分間は識別せずに前述した利用情報を収集する。そして、30 分経過ごとに過去のプロファイルと一致するかを比較する。プロファイルが一致した場合は、該当ホストは過去接続したことがあるホストであるとして、プロファイルデータを統合する。このように、30 分間プロファイルを作成し、その後過去のプロファイルと比較することでホストを特定する。ホスト特定手法を式 (1) に示す。

$$H = M \cap S \cap I \cap A \cap F \cap O \quad (1)$$

H は特定対象ホストが含まれたホスト群の集合である。そして、 M は特定対象のホストが利用したメールサーバすべてに接続したホストの集合である。ただし、特定対象ホストがどのメールサーバにも接続しなかった場合は、メールサーバを利用しないホストの集合となる。同様に S は特定対象のホストが利用した SSH サーバを利用したホスト群、 I は特定対象ホストが利用した IRC サーバと通信したホストの集合、 A は特定対象ホストが利用したすべてのアプリケーションを利用しているホストの集合、 F は特定対象ホストが発信した特徴的な通信を発信したホストの集合、 O は特定対象ホスト OS を保持しているホストの集合を指す。本論文では、式 (1) の結果、 H が 1 台しか存在しなければ、対象ホストを特定できたとする。 H が 0 台の場合は新規ホストのプロファイルとして扱う。

すでにプロファイルが作成されていた場合について述べる。図 2 に示すように、ネットワークに新しいホストが参加した場合、30 分間ユーザの挙動を取得して新規作成し

たプロファイルと既存のプロファイルを比較する。比較する情報が 1 つもとれず、10 分以内にネットワークを離れたホストのプロファイルは破棄する。そして、新規作成したプロファイルと既存プロファイルが同一であった場合はプロファイルを統合する。統合する場合、プロファイルは完全に上書きせずに差分情報をプロファイルに追加する。また、比較の際は基本的には、一番初めに作成された既存ファイルと比較対象データとし、差分情報は比較対象としない。この処置はプロファイルの増分の情報が肥大化し、すべての新規プロファイルが該当する、もしくは該当なしという結果が出るのを防止するためである。しかし、同一の内容のプロファイルが複数作成された場合や、新規作成したプロファイル情報がどの既存のプロファイルにもあてはまらない場合、候補のプロファイルが複数存在する場合には差分情報も比較対象とする。

4.3 検証

パケットヘッダ情報によるホストを識別手法の検証を 2 つの環境で実施した。表 2 に検証環境を示す。

まず、1 つ目の検証は、200 人以上の様々な企業や大学といった組織のユーザが参加するカンファレンスで行った。カンファレンスでユーザへのネットワークはすべて無線で提供されており、ユーザが保持するホストが発信したパケットヘッダ情報を収集して、特定を試みた。本手法の有効性を確認する実験をした期間は 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。次に、接続先サーバなどが類似し、本論文で定義するホストの所属す

表 3 パケットヘッダ情報によるホスト識別結果

Table 3 Result of identifying hosts by packet header information.

識別要素\ホスト	H_1	H_2	H_3	H_4	H_5	H_6
メールサーバ	M_1, M_2 M_3	M_1, M_2 M_4, M_5	M_6	M_1, M_2	M_2	M_1, M_2
SSH サーバ	S_1, S_3 S_4, S_5	S_1, S_6 S_7, S_8 S_9	S_1			
OS	MacOSX	MacOSX	Windows XP	MacOSX	Windows Vista	MacOSX
IRC Server			I_1, I_2 I_3, I_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
起動時の挙動	HTTP 通信 IMAP		HTTP	MSG	HTTP 通信	
アップデート	Evernote MacOSX	MacOSX	Windows	MacOSX	Firefox Quick Time	MacOSX Safari

H : ホスト **S** : サーバ **MSG** : Widows Live メッセンジャ

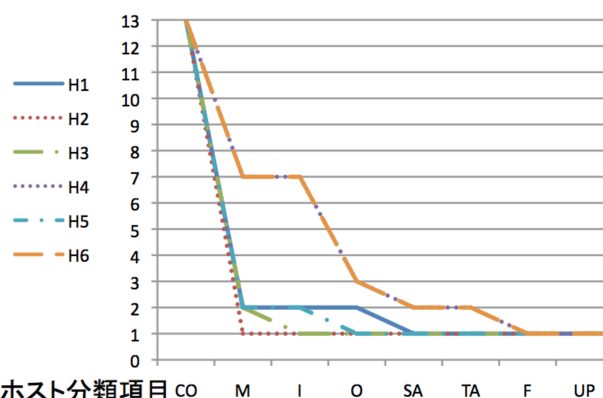
るコミュニティが同じホストが多く存在する筆者の研究室ネットワークで検証する。このネットワーク環境は、メールや Web ブラウザしか利用しないユーザも多数存在する。まず、2010 年 7 月 14 日 18 時 33 分から、2010 年 7 月 15 日 20 時 17 分のデータから特定のホストのプロファイルを作成し、2010 年 7 月 20 日 3 時 8 分から 7 月 21 日 3 時 30 分のデータを用いて、ホストの特定をした。これら 2 つのネットワークに参加するユーザは事前実験のネットワーク環境に参加するユーザと類似するため、4.2 節で述べた判別情報を利用してホストの特定を試みる。

4.3.1 カンファレンスにおける検証

200 人以上が参加するカンファレンスにおいてパケットヘッダ情報のみを利用した手法で、筆者の所属する研究会のメンバー 6 人が利用しているホストを特定できるか検証をした。対象のユーザのホスト情報は事前に収集し、カンファレンスで観測した 246 台のホストから、該当ホストを発見できたかどうかで検証する。表 3 に特定対象としたホストの情報を示す。 H_1 や H_2 は特定ホストを指し、 S_1 はある特定のサーバを指す。

まず、SSH とメールの接続先サーバを識別要素として、特定のコミュニティに分類後、ユーザのホストを特定する。SSH やメールは自身と関連するサーバを利用するため、宛先ホストからコミュニティの分類が可能である。コミュニティによってホストを分類をした結果、特定対象ホストと推測できるホストは 13 台あり、246 台から 13 台まで絞ることができた。ホスト識別結果を図 4 に示す。図 4 の Y 軸はホストの候補数である。246 台のホストからコミュニティによる分類をした後の 13 台の絞り込み結果を示している。X 軸の各項目は、識別要素をもとにホストを絞り込

候補ホスト数



ホスト分類項目 CO M I O SA TA F UP

図 4 パケットヘッダ情報によるホスト識別結果

Fig. 4 Result of identifying hosts by packet header information.

むことを示している。CO はコミュニティによる分類、M は利用メールサーバ、I は利用 IRC サーバ、O は OS、SA は OS 起動時のアプリケーション、TA は利用アプリケーション、F は特徴的な通信、UP は update 情報を識別要素とした分類を指す。

次に、高頻度で利用するメールの接続先サーバに焦点を当てる。 H_1 , H_2 は多くのメールサーバとやりと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ことが分かる。 M_1 に接続したホストは 8 台、 M_2 に接続したホストは 9 台で、 M_1 と M_2 の 2 台のみに接続したホストは 5 台、 M_3 も加えた 3 台に接続したホストは 2 台あった。 M_6 に接続したホストは 3 台あり、メールクライアント起動時に M_1 , M_2 , M_4 , M_5 の 4 台に接続を試みるホストは 1 台のみであっ

表 4 事前に作成したホストのプロファイルデータ
Table 4 Host profile data produced in advance.

識別要素\ホスト	Ha_1	Ha_2	Ha_3
メールサーバ	Ma_1, Ma_2 Ma_3, Ma_4	Ma_1, Ma_2 Ma_5	Ma_1, Ma_2, Ma_5
SSH サーバ	Sa_1, Sa_2, Sa_3	Sa_1, Sa_4, Sa_5	
IRC サーバ	Ia_1		
OS	MacOSX	MacOSX	Windows XP
特徴的な挙動	Evernote の利用 VPN 接続 Dropbox		Dropbox
起動時の挙動			Dropbox
アップデート	Evernote MacOSX	MacOSX Firefox	

H: ホスト S: サーバ

た。メールによる識別で H_2 の候補は 13 台から 1 台に絞り込むことができた。次に IRC を利用しているホストは 13 台中 1 台しか存在しなかった。OS による識別要素においては、13 台中 8 台が MacOSX、3 台が Windows であり、1 台が WindowsXP、残りは Windows Vista であった。最後に残った 2 台は Linux であることが分かった。そして、ネットワーク接続時に OS やソフトウェアのアップデート通信を取得し、FireFox を利用しているホストは 4 台であることが分かった。また、 H_1 は、起動時から定期的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。今回は、アプリケーションサーバのアドレスブロックを保持しているため、このアプリケーションが Evernote [12] であり、起動時に設定しているホストが 2 台存在することが判明した。また、 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いえる。 H_5 は発信元ポート番号として 1444 番をつねに利用している。このため、特殊なサービスを利用していると推測できる。そして、起動時に毎回 HTTP を利用したダウンロードを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。Windows はアンチウイルスソフトが導入されているため、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。実際に接続先 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ことが分かった。 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。VPN を利用しているホストは 13 台中 2 台しか観測できず、かつ MacOSX は 1 台のみであった。これにより H_6 が判明した。このように、ホストごとの特徴を抽出し分類することで、6 台中 6 台すべてを一意に特定することができた。また、機器特定の際にかかった時間は 30 分から 1 時間以内であった。機器の挙動によって特定できる時間は変化するが、本実験において 4 台のホストは 30 分間のプロファイルを作成直後に自動で

判定することができた。

4.3.2 研究室ネットワークにおける検証

作成したプロファイルで、後日ネットワークに接続したホストを特定できれば、今後特定のホストを継続して追跡できるためホストを持つ個人の追跡が可能である。そこで、筆者の所属する研究室において事前にプロファイルを作成し、ホストを分類できるか検証した。プロファイル作成期間は 2010 年 7 月 14 日 18 時 33 分から、2010 年 7 月 15 日 20 時 17 分のデータを、識別検証用のデータは 2010 年 7 月 20 日 3 時 8 分から、2010 年 7 月 21 日 3 時 30 分のデータを利用する。事前に取得したプロファイルを表 4 に示す。このプロファイルは、機器に接続して 30 分間のデータをもとに作成した。

対象ホスト Ha_1 について言及する。まず、事前に作成したプロファイルデータでは接続したメールサーバは 4 台あり、IMAP によって取得しているホストは 3 台、POP によって取得しているホストは 1 台ある。また、利用している SSH サーバは 3 台あった。利用アプリケーションは Evernote と Dropbox [13] であった。このプロファイルをもとにホストを特定する。

$$Ma_{mail1} = Ma_1 \cap Ma_2 \cap Ma_3 \cap Ma_4 \quad (2)$$

$$Sa_{ssh1} = Sa_1 \cap Sa_2 \cap Sa_3 \quad (3)$$

$$\begin{aligned} Ha_1 &= Ma_{mail1} \cap Sa_{ssh1} \cap A_1 \cap O \\ &= Sa_{ssh1} \cap Ma_{mail1} \cap A_1 \cap O \\ &= Sa_1 \cap Sa_2 \cap Sa_3 \cap Ma_1 \cap Ma_2 \\ &\quad \cap Ma_3 \cap Ma_4 \cap A_1 \cap O \end{aligned} \quad (4)$$

まず、メールサーバへのアクセスに着目してホストを分類した。検証期間中に、メールサーバ 1 に接続したホストの集合 Ma_1 は 26 台観測された。メールサーバ 2 に接続したホストの集合 Ma_2 は 35 台あり、 Ma_1 、 Ma_2 両方に接続したホストは、21 台あった。また、メールサーバ 3 に接

続したホスト群 Ma_3 は、4 台あった。 Ma_1 , Ma_2 , Ma_3 の集合すべてに含まれるホストは 3 台であった。そして、POP を利用してメールサーバ 4 に接続したホスト群 Ma_4 は 4 台であった。式 (2) にあげられるように、 Ma_1 - Ma_4 の集合すべてに含まれるホスト群 Ma_{mail1} は 3 台あった。次に、メールと同様に SSH サーバを利用したホストの分類に式 (3) を用いる。SSH サーバ 1 に接続したホストの集合 Sa_1 には 11 台のホストが、SSH サーバ 2 に接続したホストの集合 Sa_2 には 6 台、SSH サーバ 3 に接続したホストの集合 Sa_3 には 1 台のホストが観測された。そして、 Sa_1 , Sa_2 , Sa_3 のすべてに含まれるホストの集合 Sa_{ssh1} は 1 台だけであった。また、SSH と IMAP サーバに接続したホストの論理積をとったところ、1 台のみであることが判明した。そして、パケットの傾向から MacOSX であることや、Evernote を利用していることが分かった。MacOSX を利用しているホストの集合 O には 12 台、Evernote を利用しているホストの集合 A_1 には 8 台観測された。これらを式 (4) にまとめる。その結果、事前に作成した Ha_1 のプロフィールと、検証データから分類できた Ha_1 は同じホストであると推測でき、実際に Ha_1 が H_1 であることを確認した。

$$Ma_{mail2} = Ma_1 \cap Ma_2 \cap Ma_5 \quad (5)$$

$$Sa_{ssh2} = Sa_1 \quad (6)$$

$$\begin{aligned} Ha_{2-1}, Ha_{2-2} &= Ma_{mail2} \cap Sa_{ssh2} \cap A_{stapp} \cap O \\ &= Sa_1 \cap Ma_1 \cap Ma_2 \cap Ma_5 \\ &\quad \cap A_{stapp} \cap O \end{aligned} \quad (7)$$

Ha_2 について述べる。事前に作成したプロフィールデータによると、 Ha_2 はメールサーバは Ma_1 , Ma_2 と Ma_5 の各集合に含まれる。また、SSH サーバでは Sa_1 , Sa_4 , Sa_5 の集合に含まれていることが分かった。このプロフィールをもとにホストを特定する。まず、式 (5) を用いて、メールサーバによる分類をする。 Ma_1 , Ma_2 は前述したとおりであるが、メールサーバ 5 に接続したホストの集合 Ma_5 には 25 台のホストがあった。 Ma_1 , Ma_2 , Ma_5 の集合すべてに含まれるホスト群 Ma_{mail2} は 7 台観測された。SSH サーバにおいて、 Sa_1 の集合には 11 台のホストが確認されたが、 Sa_4 , Sa_5 はどのホストも含まれなかった。そこで、式 (6) にあげるように、 Sa_{ssh2} と Sa_1 は同じ集合とする。そして、 Ma_{mail2} と Sa_{ssh2} の両方に含まれるホストの集合には 2 台ホストがあった。また、起動時に Dropbox を利用しているホスト A_{stapp} は 3 台あった。そして、ホストが MacOSX である集合 O を付け加えて、式 (7) を用いたが、2 台のホスト Ha_{2-1} , Ha_{2-2} のどちらが Ha_2 であるか特定できなかった。その原因を調査したところ、 Ha_{2-1} , Ha_{2-2} は 1 台のホストの無線と有線のインタフェースであり、1 つのホストからパケットが発信していることが判明

した。そのため、ホストの特定はできたといえる。

$$H_3 = Ma_{mail2} \cap A_2 \cap O_{winxp} \quad (8)$$

H_3 について述べる。 H_3 は SSH を利用せず、メールや HTTP 通信を利用するホストである。また、発信元ポート番号から Windows XP であり、Dropbox を利用している。それらのプロフィールデータをふまえて、検証データから H_3 を特定する。メールサーバによる識別においては、 H_2 の場合と同じく式 (5) を用いて 7 台のホストが該当することが分かった。OS に関しては Windows XP を利用しているホストの集合 O_{winxp} には 5 台のホストが含まれる。また、Dropbox を利用しているホストの集合 A_2 には 2 台のホストが該当した。これらを式 (8) にまとめる。Windows XP を利用しており、かつ Dropbox を利用し、 Ma_{mail2} に含まれるホストは 1 台のみであり、これが対象ホスト H_3 であることを確認した。これにより、複数の同じコミュニティのユーザが多く存在する環境において、3 台の特定対象ホストすべてを特定できた。

また、機器特定の際にかかった時間はカンファレンスにおける検証と同じく 30 分から 1 時間以内であった。本検証において 2 台のホストは 30 分間のプロフィールを作成直後に判定することができた。

5. ホストとユーザを結びつける手法

ホスト利用者を判別するにあたり、誰がどの機器を保有しているかといったユーザ情報はネットワーク管理者や、ユーザに応じたサービスを提供する側にとっては重要な情報である。特に、近年スマートフォンやノートブックの台頭によって、1 人で複数の機器を持つユーザが増加しているため、誰がどのホストを何台持っているかの情報を取得することによって、より効率的にサービス提供やネットワーク管理が可能である。しかし、前述したパケットヘッダ情報による手法では、ホストの特定、追跡が可能であるが、ユーザの特定までは不可能である。ホストとホスト利用者を結びつける方法に、ユーザにホストの機器情報と本人の名前といった識別 ID を登録するアカウント登録がある。しかし、アカウント登録はユーザに操作を要求するため不特定多数のユーザが利用する場合や、管理の緩いネットワークにおいて、ユーザにアカウント登録を義務付けられない場合がある。また、アカウント登録は容易に詐称が可能であるため、完全にユーザを特定する場合には適さない場合が多い。

そこで、視認情報などの対象ユーザのネットワーク在席情報をもとにホストとホスト利用者を結びつける手法を提案する。本手法は、GPS などの専用デバイスの利用といった制限をなくすとともに、対象ホストの利用者および、1 人で複数の機器を持つユーザを特定することを目的とする。本手法は、パケットヘッダ情報による識別でホストを一意

に特定できた場合を想定し、特定できたホストとユーザを結びつける。具体的な方法として、ターゲットとするホスト所有者が特定の時間、ネットワークを利用している情報、もしくは利用していない情報をホストの情報と組み合わせることでホスト所有者を特定する。ターゲットとするホスト所有者の在席情報を確認する方法は複数存在する。たとえば、情報収集者が、直接ターゲットのホスト利用者がネットワークを利用しているかを視認や、会議の出欠名簿などのリストから判断する、同じ場所でターゲットのホスト利用者と一緒にいたユーザがその情報を伝達、もしくは監視カメラなどにターゲットが映っている情報などを利用することがあげられる。今回は、前述したホスト所有者がネットワークの利用の有無の確認に、情報収集者による視認も用いた手法を利用する。対象のユーザが、在席した時間、退席した時間とパケットのヘッダ情報によるホスト識別情報の論理積をとることにより、ホスト利用者の特定が可能である。これによって、アカウント登録と比べて、ユーザがMACアドレスだけでなくホスト自体も変更した場合や、新しくネットワークに参加したデバイスをネットワークに持ち込んだ場合にも発見することが可能である。しかし、筆者が提案している視認によるユーザの識別手法 [14] では、所有者の在席時間のみでは確実に絞ることができなかった。そこで、本論文では対象ユーザの不在情報を追加することにより、ホストとホスト利用者の結びつけの手法を改良し、精度を向上した手法を提案する。

5.1 判別手法

図 5 に本手法の概要を示す。はじめに、観測者 1 がユーザ X のホストと同じ場所で行動した場合、観測者 1 はユーザ X が場所 A にいるという情報をホスト A を利用して本システムに通知する。通知する内容は、物理的な場所、対象ユーザの名前、時間の 3 つが含まれる。本システムはその情報を収集し、入力があったタイミングの前後 5 分間に観測されたホストの中に、ユーザ X のホストが含まれているとする。次に、後日観測者 2 がユーザ X のホストと同じ場所で行動した場合も同様にホスト C を利用して本シ

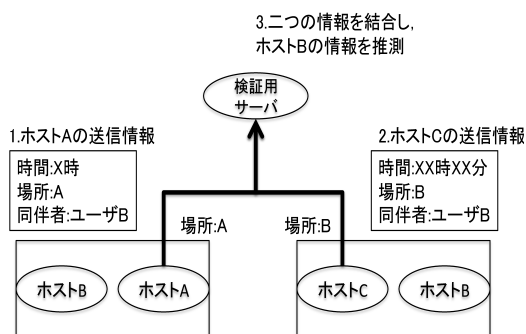


図 5 ホスト情報を推測する手法

Fig. 5 Technique for inferring hosts.

テムに通知することで、ユーザ X のホストを過去のデータと照らし合わせて絞り込むことができる。この動作をホストが 1 つになるまで繰り返せば、ユーザ X の持つホスト B を特定できる。絞り込みによるホストを特定する方法を式 (9) に示す。H が対象とするホストの集合、 TH_n は n 回目のターゲットユーザが存在する視認情報を入力した時間、前後 5 分間に検出された H の候補となるホストの集合である。また、本人が存在しなかったという情報も有用な識別要素となる。ターゲットユーザが存在しないホスト群の NH とし、m 回目のターゲットユーザが存在しないホスト群を NH_m とする。これによって、前述した常時稼動している機器があった場合でも、正確に情報を取得できる。これら、 TH と NH_m の論理積により H が 1 台しか存在しなければ、ホストとホスト利用者が結びつけることができたといえる。

$$H = TH_1 \cap \dots \cap TH_n \cap \overline{NH_1} \cap \dots \cap \overline{NH_m} \quad (9)$$

本手法はホスト利用者がネットワークを利用している情報をもとに、ユーザを特定するが、常時稼動する機器を除外することでより精度を向上させることができる。たとえば、ホスト利用者が長時間同じネットワークを利用しているケースがあげられる。その場合、ユーザがネットワークに存在しない場合の情報を取得できず、ホスト利用者がネットワークを利用している情報のみであった場合、式 (9) の手法では、常時稼動するホストも候補として組み込まれてしまう。そのため、可能であれば常時稼動しているホストやサーバを、事前に本手法で収集し、除外することで精度を向上させることが期待できる。

5.2 視認によるホスト利用者特定手法の検証

実際に本手法でホスト所有者を特定できるか、筆者の所属する研究室で検証した。筆者の所属する研究室は 100 人程度のユーザが在籍し、ネットワーク上では 200 台を超えるホストが観測できる。また、研究室には誰か 1 人は在席しており、数日間連続して宿泊するユーザも存在する。このような環境において、無作為に選出した 3 台のホスト情報から、ホスト所有者が特定できるか検証した。検証方法はそのユーザが所属するネットワークに出現しているのを筆者が視認したうえで情報を入力する。

しかし、検証環境において、長期間泊まりこむホスト利用者も存在する特殊な場合も考えられるため、5.1 節で述べたように、常時稼動する機器と判別が困難となる例外が発生する可能性がある。その例外に対応するため、常時稼動しているホストをリストから除外する必要がある。そこで、3 日前からデータを収集し、取得した 363 台のホストから常時稼動しているホスト 66 台のリストを作成、除外した。そして、各対象ユーザの在席情報と退席情報を交互に入力した。図 6 に常時稼動ホストを除外した結果を示す。

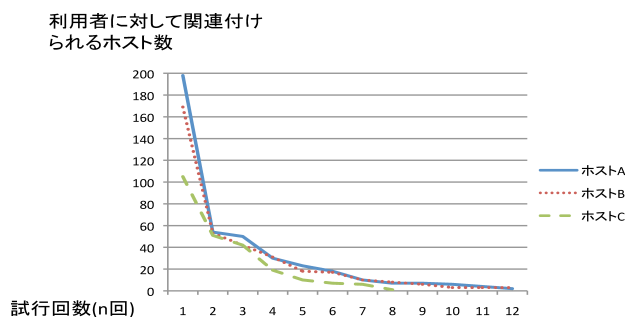


図 6 ホスト識別手法結果

Fig. 6 Result of identifying hosts.

ホスト A, B は 12 回で 2 台, ホスト C は 8 回の試行で 1 台に絞り込むことができた. ホスト C に関しては, ホストとホスト利用者を確実に結びつけることができた. ホスト A, B は 1 台まで絞り込みが困難であった理由は, 1 人でノートパソコンとスマートフォンの 2 つのデバイスを所持するユーザであったため, 一意に特定ができなかったのが原因であった. このように複数回以上の施行で 1 台にまで絞りきれなかった場合は, 対象の 2 台の機器がネットワークに出現したすべての時間を参照し, その時間が一致するかを確かめる. 今回の場合は 2 つの機器がネットワークに出現した時間がまったく同一であったため, ユーザが複数所持しているホストであると判定した. これらの結果から在席情報と退席情報を取得した視認によるホスト利用者特定する手法は有効であるといえる.

6. 考察と今後の課題

パケットのヘッダ情報をもとにホストの識別をする手法を検証した結果, メールサーバや SSH サーバに接続したという情報は, 大学に所属するユーザや企業に所属するユーザが多く存在する環境において重要な識別要素となることが判明した. この識別要素を利用することでホストをある程度, もしくは完全に特定することができる. 特にメールサーバの接続に関しては, メールクライアント利用時に事前に設定した複数のメールサーバ群との通信が必ず発生するため, ホストの特定では重要な識別要素であるといえる. その後, アプリケーションや OS 情報を利用してホストの特定をする. これによって, パケットのヘッダ情報のみからホストを絞り込むことや, 複数インタフェースからパケットを発信する 1 台のホストを特定できることを示した. 実際に事前にプロファイルを作成した 3 台のホストに関して, 本人拒否率は 100% であり, 他人受入率は 0% という検証結果が出た. 本提案手法によって, ユーザアカウント登録や専用デバイスの利用をユーザに強いることなく, ユーザのホストを検出することが可能であることを示した. しかし, 重要な識別要素となった SSH や IRC は, メールや Web ブラウジングのみのユーザも存在するため, SSH サーバを利用しないようなユーザの特定も必要である. そ

のような一般的なユーザは, カンファレンスでの検証では H_3 , H_4 , 研究室ネットワークの検証では H_3 を特定したことで可能であることを示した.

本論文では SSH やメールの接続サーバを識別要素としてあげた. これは対象のネットワークの参加者が大学や企業に所属するユーザが多く存在するカンファレンスであり, SSH やメールが多く利用されている環境であることが事前検証で判明したためである. 特定のユーザが情報を利用するデバイスは, ユーザの利用形態の特徴によって発信する情報に特徴的なトラフィックパターンが現れるため, ユーザが参加するネットワークによっては取得する情報や識別要素も異なる. そのため, 識別要素として利用できるプロトコルは企業やカンファレンス, 一般的な家庭ネットワークなど環境によって変化する. たとえば, 日本の 2008 年度のインターネットのトラフィックの傾向 [15] によると SSH は全トラフィックの 0.1% 程度しか利用されておらず, 一般的なネットワークでは SSH を識別要素として利用は困難である. しかし, インターネット白書 2009 [16] によるとリモートログインを許可している企業は約 40% 以上にのぼるため, 企業のネットワークにおいては SSH の割合は識別要素として利用できる可能性が高いと推測できる. また, ISP などが提供する一般的なネットワークにおいて本手法を適応する場合は, ホストのインストール済みのアプリケーションに着目することで識別が可能であると推測できる. たとえば, スタートアップ時のホストの挙動や, メッセンジャや, スタートアップに登録されているアプリケーション, OS や, メールクライアントの使用の有無などを識別要素として利用できる. このように, ユーザの振舞いによって取得可能な情報は変化するため, ネットワークに依拠して, 識別要素を利用することで様々なネットワークでホストの識別が可能である.

本手法が対象とするアプリケーションの追加や, 特徴の抽出をさらに細かく検討する必要がある. また, サービスを特定するために特定サービスの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に, 本システムの識別要素として利用している情報がとれないことが考えられる. そのため, 今後サービスごとにトラフィック量や通信パターンからサービスを推測する必要がある. このように今後, 様々なホストの特定ができるように, より多くの識別要素を検討する.

ユーザのネットワーク在席情報を利用した識別手法において, ホスト所有者の特定が可能であることを示した. 今回は情報収集者の視認による識別情報を利用したが, それ以外の情報も利用可能である. たとえば, 管理者自身がリアルタイムに情報を入力することなく, 一般ユーザから対象ユーザの情報を得ることや, 会議の参加者名簿や会議のログなどと連動させることにより, 管理コストの低減が可能である. しかし, アカウント登録手法と比較すると, ユー

ザ全員にアカウント登録の義務付けが困難なネットワークなどでの利用や、MAC アドレスなどを詐称された場合でもホストやホスト利用者を特定できる反面、ネットワーク管理者の管理コストはネットワーク参加者の数に比例する。そのため、ネットワークの環境や目的によってアカウント登録手法や本手法を組み合わせることで、より効率的にホストとホスト利用者を結びつけることができる。

本検証環境は、必ず1人のユーザが存在する環境で行った。そのため、多くのユーザが同時に在席、退席する企業のような組織のネットワークとは異なる。しかし、企業や他のネットワークにおいても、ユーザの在席情報、退席情報は用事や仕事などで異なるため、検出できると推測できるが、実際に検証する必要がある。今後は、管理者以外のユーザが管理者に対象ユーザの在席情報を伝えられる基盤を提供する。

本論文では、パケットヘッダ情報と、ユーザのネットワーク在席情報を利用することにより、ユーザの特定が可能であることを示した。本手法によって、サービス提供者はユーザに意識させることなくサービスを提供できる反面、ユーザプライバシーに配慮する必要がある。本手法により、パケットヘッダ情報のみでホストの追跡が可能であるため、トラフィックデータの収集や譲渡の際にはユーザの同意を得る、オプトイン形式にするなどサービス提供者はユーザプライバシーの配慮が求められる。

7. 結論

本論文の目的は、個人情報と定義されていない情報を効果的に活用することによって、ホストやホスト利用者を特定し、利用者を追跡することである。これによって、サービス提供者はユーザに合わせたサービスの提供ができ、ユーザも意識せずにサービスを利用することができる。

パケットのヘッダ情報を利用する手法では、MAC アドレスを除くパケットのヘッダ情報から各ホストのプロファイルをもとに、該当するホストを特定した。ネットワークに参加するユーザによって、インターネットの利用形態や特徴的なトラフィックは大きく変化する。そのため、今回は1つの利用形態として企業や大学のユーザが多く参加するネットワークにおいて、デバイスの特徴的なトラフィックから識別要素を取得し、それをもとにホストを識別した。その結果、特徴的な通信をするホストだけでなく、Web ブラウジングとメールを主に利用するホストも識別が可能であることや、本人拒否率は100%であり、他人受入率は0%という検証結果が得られた。そして、1人で複数のホストを利用している場合も検知可能であることも示した。また、今回は研究ネットワークから識別情報を取得したが、他のネットワークにおいても同様にトラフィックから特徴的な通信を取得することで、識別要素を抽出し、ホストを識別することが可能である。

以上の結果をふまえ、パケットヘッダ情報と特定対象ユーザのネットワーク在席情報を利用することで、ユーザを特定可能であることを示した。今回利用した情報はパケットヘッダ情報とユーザのネットワーク在席情報であるが、利用可能な情報は他にも数多くある。そのため、さらに調査、検証する範囲を広げ、ユーザを特定できる可能性がある情報を収集する。これにより、サービス提供者やネットワーク管理者はユーザに特定の操作やデバイスの制約を受けることなく、よりユーザに特化したサービスの提供やネットワークに参加しているホストや利用者の状態が容易に把握できるようになる。

参考文献

- [1] Karagiannis, T., Papagiannaki, K. and Faloutsos, M.: BLINC: Multilevel Traffic Classification in the Dark, *Proc. ACM SIGCOMM*, pp.229–240 (2005).
- [2] Montigny-Leboeuf, A.D.: Passive Network Discovery for Real Time Situation Awareness, *Proc. RTO Information Systems Technology Panel (IST) Symposium on Adaptive Defence in Unclassified Networks*, pp.288–300 (2004).
- [3] 和泉勇治, 佐藤 剛, 田中和之: パケットペイロード長の発生確率に基づいたネットワークアプリケーション識別 (BCI/BMI その周辺, 一般), 電子情報通信学会技術研究報告 NC, ニューロコンピューティング, Vol.110, No.295, pp.7–12 (2010).
- [4] Krishnamurthy, B. and Wills, C.: On the leakage of personally identifiable information via online social networks, *Proc. 2nd ACM Workshop on Online Social Networks*, pp.7–12, ACM (2009).
- [5] 松尾 豊, 友部博教, 橋田浩一, 中島秀之, 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, Vol.20, pp.46–56 (2005).
- [6] 本村憲史, 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, 電子情報通信学会技術研究報告 OFS, オフィスシステム, Vol.98, No.75, pp.29–36 (1998).
- [7] Electronic Frontier Foundation: How unique is your browser?, *Privacy Enhancing Technologies Symposium* (2010).
- [8] 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 (2008).
- [9] 総務省: 利用者視点を踏まえた ICT サービスに係る諸問題に関する研究会第二次提言 (2010), 入手先 (http://www.soumu.go.jp/main_content/000062003.pdf).
- [10] Zalewski, M. and Passive, O.: Fingerprinting Tool (2010), available from (<http://lcamtuf.coredump.cx/p0f.shtml>).
- [11] dokodoko.jp: dokodoko.jp (2011), available from (<http://dokodoko.jp/>).
- [12] Evernote Corporation: Welcome to your notable world (2010), available from (<http://www.evernote.com/>).
- [13] Dropbox: Dropbox (2010), available from (<https://www.dropbox.com/>).
- [14] 上原雄貴, 水谷正慶, 武田圭史, 村井 純: デバイス情報収集によるユーザ追跡システムの試作, コンピュータセキュリティシンポジウム 2010, Vol.2010, pp.837–842 (2010).
- [15] Cho, K., Fukuda, K., Esaki, H. and Kato, A.: Ob-

serving slow crustal movement in residential user traffic, *Proc. 2008 ACM CoNEXT Conference, CoNEXT '08*, pp.12:1–12:12, ACM, New York, NY, USA (2008).

- [16] 財団法人インターネット協会：インターネット白書 2009, 株式会社インプレス R&D (2009).



上原 雄貴

昭和 62 年生。平成 22 年慶應義塾大学総合政策学部卒業。同年同大学大学院政策メディア・研究科前期博士課程入学。現在に至る。ユーザ追跡やプライバシーの分野を中心に情報セキュリティの研究に従事。



水谷 正慶 (正会員)

昭和 58 年生。平成 18 年慶應義塾大学環境情報学部卒業。平成 20 年同大学大学院政策・メディア研究科前期博士課程修了。平成 22 年同大学院政策・メディア研究科後期博士課程修了。博士（政策・メディア）。同大学訪問研究員を経て、平成 23 年 4 月より IBM 東京基礎研究所。現在に至る。ネットワーク侵入検知、不正プログラム対策分野を中心に情報セキュリティの研究に従事。



武田 圭史

慶應義塾大学大学院政策・メディア研究科後期博士課程修了。博士（政策・メディア）。防衛庁航空自衛隊、アクセシチュア株式会社、カーネギーメロン大学情報ネットワーク研究所、カーネギーメロン大学日本校教授、慶應義塾大学大学院政策・メディア研究科教授を経て、平成 21 年 4 月より慶應義塾大学環境情報学部教授。現在に至る。ネットワーク侵入検知、事件対応、不正プログラム対策分野を中心に情報セキュリティの研究に従事。兵庫県参与（情報セキュリティ担当）。



村井 純 (正会員)

昭和 54 年慶應義塾大学工学部数理工学科卒業。昭和 56 年同大学大学院修士課程修了。昭和 62 年同大学院博士（工学）。昭和 59 年東京工業大学助手。昭和 62 年東京大学助手。平成 2 年慶應義塾大学助教授を経て、平成 9 年より同大学教授。平成 11 年より平成 17 年まで慶應義塾大学 SFC 研究所所長。平成 15 年より Auto-ID ラボジャパン所長。平成 17 年より平成 21 年まで慶應義塾常任理事兼慶應義塾大学環境情報学部教授。平成 21 年慶應義塾大学環境情報学部教授兼学部長。