マジックミラーの裏側で：企業監視テクノロジーの詳細

ベネット・サイファーズ,ジェニー・ゲブハート(電子フロンティア財団)`

2019年12月2日

目次

1. 前書き[1]

追跡者(トラッカー)は、今日のインターネットのほぼすべての場所に、つまり現代生活のほぼすべての場所に隠れています。 平均的な Web ページは、多数のサードパーティ[2]とデータを共有しています。 平均的なモバイルアプリでも同じことが行われ、多くのアプリは、使用していないときでも場所や通話記録などの極めて機密性の高い情報を収集します。トラッキング[3]は物理的な世界にまで及びます。ショッピングセンターでは、自動ナンバープレート読み取り装置を利用して駐車場を通行する自動車の往来を追跡し、そのデータを法執行機関と共有します。企業、コンサート主催者、政治キャンペーンでも、Bluetooth や WiFi 電波を利用して、地域の人々の受動的な監視(passive monitoring)を実行しています。小売店は、顔認証を利用して顧客を識別し、盗難のスクリーニングを行い、ターゲット広告を配信しています。

この監視の背後にあるハイテク企業、データブローカー、広告主、そしてこれらを推進している技術は、平均的なユーザーからはほとんど見えません。こうした企業はマジックミラーが設置されている部屋を作り出したわけです。部屋の内側からは、アプリ、ウェブページ、広告、ソーシャルメディアに写し出された自分しか見えません。しかし、鏡の後ろで、トラッカーはあなたのほとんどすべての振舞いをじっと記録しています。こうしたトラッカーは、全知ではありませんが、広範にどこにでも存在します。彼らが収集して導き出すデータは完全ではありませんが、非常に機密性の高いものです。

この白書では、「サードパーティ」企業のトラッキング、つまりユーザーがやり取りするつもりのない企業による個人情報の収集に焦点を当てます。サードパーティの追跡の背後

1 ここに訳出したのは、Bennett Cyphers、Behind the One-Way Mirror: A Deep Dive Into the Technology of Corporate Surveillance、Electronic Frontier Foundation, December2, 2019, https://www.eff.org/wp/behind-the-one-way-mirror#Data-brokers です。このレポートは企業がどのように私たちの個人情報を収集しているのかについて、その裏側の仕組みを詳細に解説しています。言語では One-way mirror となっていますが、翻訳では「マジックミラー」と訳しましたが、まさに私たちからは見えないけれども鏡の後ろからじっと私たちの行動や心理を観察している仕組みがあるということをこのことばで端的に表現していると思います。レポートの対象は米国ですが、ここで述べられている多くのことは、法制度は違っても日本でも共通していえることと思います。訳注を付して、IT やビジネス用語の解説を補足しましたが、大半はネットの IT 用語辞典などからの引用あるいは要約です。辞典横断検索 Metapedia http://metapedia.jp で検索しました。また、翻訳に際しては、google tranlate、babel fish などの機械翻訳や英辞郎も参照しました。(小倉利丸 JCA-NET)

2 訳注： サードパーティ 二者が関係する契約事などについてよく用いられる表現で、当事者については一方を「ファーストパーティ」(first party)、その相手方を「セカンドパーティ」(second party)という。当事者が三者以上に渡る場合も、慣用的に当事者以外の関係者をサードパーティということが多い。
　ビジネスの用語としては、ある製品やサービスなどについて、その開発・販売元以外で、それに関連する、あるいは対応する製品を開発・販売する事業者のことをサードパーティということが多い。この場合、大元の製品の開発元がファーストパーティ、その顧客や利用者がセカンドパーティとなる。
(http://e-words.jp/w/%E3%82%B5%E3%83%BC%E3%83%89%E3%83%91%E3%83%BC%E3%83%86%E3%82%A3.html)

3 訳注：トラッキング 人の行動やシステムの挙動、データの推移などの情報を継続的に収集、監視すること。インターネット上で Web サイトの来訪者やオンラインサービスの利用者を識別し、訪問履歴やサイト内での行動を記録・追跡することをトラッキングという。Web サーバと Web ブラウザの間で永続的にデータを共有する HTTP Cookie (クッキー) と呼ばれる仕組みを用いる。
(http://e-words.jp/w/%E3%83%88%E3%83%A9%E3%83%83%E3%82%AD%E3%83%B3%E3%82%B0.html)

にある技術的な方法とビジネス慣行に光を当てます。ジャーナリスト、政策立案者、および関心を持つ消費者のために、サードパーティによるトラッキングの基礎を分かりやすく説明し、問題の範囲を説明し、現状に対抗するユーザーと法的手段を提案しようと思います。

パート 1(本翻訳では第 3 章)では、「識別子」[4]、つまりトラッカーが Web、モバイルデバイス、および物理的な世界で、誰が誰なのかかをトラッキングするために利用する情報を分類します。識別子により、トラッカーは行動データを実際の人にリンクすることができるようになります。

パート 2(本翻訳では第 4 章) では、企業がこれらの識別子やその他の情報を収集するために使用する手法について説明します。 また、巨大なトラッカーたちが監視ネットワークの構築を支援するように他の企業をどのように説得するかについても述べます。

パート 3(本翻訳では第 5 章)では 、異種のアクター[5]が互いに情報を共有する方法とその理由について詳しく説明します。すべてのトラッカーがあらゆる種類のトラッキングを行うわけではありません。その代わりに、細分化された企業の Web は、様々なコンテキストに沿ってデータを収集し、特定の目標を達成するためにデータを共有したり販売したりします。

最後に、 第 4 部(本翻訳では第 5 章)では、消費者と政策立案者が反撃するために実行できるアクションを提示します。まず、消費者はツールや行動を変えることによって、デバイスでのトラッキングをブロックすることができます。政策立案者は、サードパーティの追跡を抑制する包括的なプライバシー法を採用すべきであることを提示します。

2. 序論

2.1. ファーストパーティ vs サードパーティのトラッキング

インターネット上の巨大企業は、人々がそのサービスを利用する際に大量のデータを収集します。Facebook は、あなたの友人が誰なのか、あなたが何に「いいね」をしたのか、あなたがニュースフィードで読んだコンテンツの種類を知っています。Google は、Google マップでナビゲートしているときに、検索内容と移動先を認識します。アマゾン

4 訳注：識別子 ある実体を他から区別して指し示すのに用いられる情報で、単一の文字列や数字列などが用いられることが多いが、複数の情報を組み合わせて用いることもある。極めて幅広い概念で、システムの種類や取り扱う対象の種類によって多種多様な識別子があり、命名規則や利用できる文字の種類などもそれぞれ異なる。
コンピュータによって割り当てられる識別子は重複や衝突がないように一定の手順で機械的に生成して一意に定めることが多いが、現実世界で用いられる人名などの識別子は必ずしも一意ではない。情報システムで取り扱う際には通し番号など独自に識別子を生成して割り当てたり、他の情報と組み合わせるなどの工夫が必要になることが多い。
識別子は多岐にわたるため例示しきれないが、例えば、OS における識別子には、ファイル名やフォルダ名、ドライブ名、ユーザー名、アカウント名、インターフェース名、プロセス ID、UUID、GUID などがある。プログラミング言語やコンピュータプログラムにおける識別子には、変数名や関数名、クラス名、インスタンス名、メソッド名、プロパティ名、ラベルなどがある。通信・ネットワークシステムにおける識別子には、IP アドレスやドメイン名、ホスト名、ポート番号、MAC アドレス、SSID、メールアドレス、URL/URI などがある。 現実世界で人や物、組織、場所などを指し示す識別子もたくさんあり、氏名や法人名、住所や所在地、建物名、部屋番号、電話番号、自動車登録番号（ナンバープレート）、緯度、経度など様々な情報が識別子として利用されうる。
「ID」という語は“identification”の略とされることもあり、識別や特定、同定などの動作・作業のことや、（人物の）身分証明（書）などの意味で用いられる。(http://e-words.jp/w/%E8%AD%98%E5%88%A5%E5%AD%90.html)
5 訳注： アクター 「〔活動などの〕当事者、参加者、動作主体」(英辞郎)などと訳されるが、ここでは、トラッカーとして活動する企業や広告主、Google、Amazon、Facebook といったユーザにコミュニケーションの場を提供する企業などを指す。以下はその解説の一部 「 広告が狙いどおりの利用者に表示されているかを確認する。新しい潜在顧客を見つけたり、ウェブサイトで特定のページにアクセスした人や望ましいアクションをとった利用者にリーチしたりできます。 販売を促進する。自動入札機能を利用すると、求めているアクション(商品の購入など)を実行する可能性が高い利用者にリーチできます。 広告の効果を測定する。利用者が広告を見たときのアクションを測定することで、広告の効果についての理解を深めることができます。

は、あなた買い物の目的や何を買ったのかを知っています。

これらの企業が自社の製品およびサービスを通じて収集するデータは「ファーストパーティデータ」と呼ばれます。この情報は非常に機密性が高く、誤って取り扱われた多くの実績があります。ファーストパーティのデータは、暗黙または明示的に契約の一部に含まれて収集される場合があります。つまり、当社のサービスを使用することを選択し、お客様は当社が収集するデータを使用することに同意するものとします、という契約です。多くのユーザーは、多くの無料サービスが、たとえ気にらなくても、製品であることを理解するようになっています。

ただし、企業は 、そのサービスを使用していない人に関する個人情報も同様に収集しています。たとえば、Facebook は目に見えない「コンバージョンピクセル」[6]を使用して、他のウェブサイトやアプリのユーザーに関する情報を収集します。同様に、Google は位置データを使用してユーザーの実店舗への訪問を追跡します。そして、他の何千ものデータブローカー、広告主、トラッカーが、日々の私たちのウェブ閲覧やデバイス利用[7]の背後に潜んでいます。これは「サードパーティによるトラッキング」として知られています。サードパーティのトラッキングは訓練された目がなければ識別することが極めて困難で、これを完全に回避することはほぼ不可能です。

2.2. 彼らは何を知っているのか？

多くの消費者は、自分が使っているデバイスに対する最もはっきりわかるプライバシー侵害の可能性についてはよく知っています。すべてのスマートフォンはポケットサイズの GPS トラッカーであり、インターネットを介して見知らぬ人たちに居場所を常時送信(ブロードキャスト[8])します。カメラとマイクを備えたインターネット接続デバイス[9]は、密かな盗聴装置に転換される固有のリスクを抱えています。そして、リスクは現実のものでもあるのです。これまでにも、位置データは著しく乱用されてきました。Amazon と Google はどちらも、従業員は、Alexa や Home といった個人宅のリスニングデバイスで録音された音声を聞くことが許されています。また、ユーザを写すように設置されているラップトップのカメラは、学校や自宅で生徒をスパイするために利用されています。

しかし、これらのよく知られている監視の回路は、最も一般的なものでもなければ、必ずしも最もプライバシーを脅かすものだともいえません。私たちは、目を覚ましている時間の多くを、インターネットに接続されたカメラの視線のなかで暮していますが、これらがユーザーの自覚的な意図なしに何かを録画することは極めてまれです。連邦 および州の盗聴に関する法 に違反しないように、ハイテク企業は通常、ユーザーの会話を密かに聞くことも控えています。以下で示すように、トラッカーは、それほど劇的とはいえないような数千のデータソースから十分な情報を取得します。実際にある不安な事態というのは、Facebook があなたの電話を通してあなたを盗み聞きすることにあるのではなく、そんなことをする必要がない(そうしなくても分る) というところにあります。

私たちのプライバシーに対する最も一般的な脅威は、私たちがどのように生活しているかに関するデータが、徐々にですが恒常的に容赦なく蓄積されているという点にあります。これには、閲覧履歴、アプリの使用状況、購入、地理データなどが含まれます。これらの些細な部分は、膨大なデータと組み合わすことができてしまうのです。トラッカーは、ク

6 訳注： コンバージョンピクセル Facebook の場合は「Facebook ピクセル」と呼んでいる。Facebook ピクセルを設定すると、ウェブサイトで利用者がアクションを起こした場合にピクセルが実行されます。アクションには、例えばショッピングカートへのアイテムの追加や、購入があります。ピクセルはこれらのアクション(イベント)を受信します。アクションはイベントマネージャの Facebook ピクセルページで表示できます。このページで、顧客のアクションを確認することができます。今後、これらの顧客に Facebook 広告で再度リーチするための選択肢も用意されています。」

7 訳注：デバイス コンピューターに接続して使うあらゆるハードウェアのこと。マウス、キーボード、プリンター、ディスクドライブ、ネットワークアダプターなど、あらゆる装置を指す。CPU や電源など、内蔵されている部品もデバイスに含まれる。パソコンで新しいデバイスを使うには、デバイスドライバーというソフトウェアをインストールする。(https://kotobank.jp/word/%E3%83%87%E3%83%90%E3%82%A4%E3%82%B9-6375) スマホや IoT のようなインターネットに接続された「モノ」もデバイスである。

8 訳注： ブロードキャスト ネットワークに参加するすべての機器に同時に信号やデータを送信すること。

9 訳注： スマートフォンやカメラ付きのパソコンなど。

リック、インプレッション[10]、タップ、および動きに関するデータを広範な行動プロファイルにまとめ、政治的所属、宗教的信念、性的アイデンティティや活動、人種とエスニシティ、教育レベル、所得水準、購買習慣、身体的および精神的健康を明らかにすることができます。

膨大な個人情報が収集されるにもかかわらず、トラッキング企業はこのデータを使用してしばしば不正確だったり誤った結論を導きます。行動ターゲティング広告とは、ユーザーの行動に関するデータを使用して、何が好きか、どのような考え方か、購入可能性のあるものは何かを予測することで、これがサードパーティのトラッキング産業の多くにとっての動機付けになります。行動ターゲティングの広告主は正確な情報にアクセスできる場合がありますが、大まかな一般化と「ないよりマシ」といった程度の統計的推測を扱うこともよくあります。ユーザーは、驚くほど正確な広告と笑うしかないような的外れな広告の両方がウェブ上にあらわれるのを見ることになります。マーケティング業界全体で、トラッカーはペタバイト[11]単位の個人データを使用してデジタルな占いをやっているのです。トラッカーの推論が正しいかどうかに関係なく、収集されたデータはプライバシーを著しく侵害しており、そのデータに基づいて行う決定は具体的な損害を引き起こす可能性を持っています。

3. いずれにせよ誰のデータか：トラッカーはどのようにデータと人を結び付けるのか？

ほとんどのサードパーティのトラッキングは、実在の人物のプロファイルを作成するように設計されています。つまり、トラッカーが情報を収集するたびに、その情報を特定の人物と結び付けて利用できるような識別子が必要になります。トラッカーは、これを間接的な形で行うことがあります。つまり、収集されたデータを特定のデバイスまたはブラウザと相関させ、これが後に個人や家庭のような小さなグループと関連付けられる場合があるのです。

誰が誰であるかを追跡し続けるためには、トラッカーはユニーク[12]で永続的で利用可能な識別子が必要です。つまり、トラッカーは、（1）ユーザーまたはデバイスのみを示す情報、（2）変更のない情報、および（3）簡単にアクセスできる情報を探します。いくつかの潜在的な識別子はこれら3つの要件すべてに適合しますが、トラッカーはこれら3つのうち2つだけをチェックする識別子も使用し続けることがきます。トラッカーは、複数の強力とはいえない識別子を組み合わせて、強力な識別子を作成することができます。

上の3つすべてをチェックする識別子は、名前、電子メール、または電話番号です。「af64a09c2」や「921972136.1561665654」といったものが、トラッカーによってあなたに与えられる「名前」かもしれません。トラッカーにとって最も重要なのは、識別子があなたを、あなただけを示しているということにあります。時間が経つにつれて、

10 訳注： インプレッション Webサイトに掲載される広告の効果を計る指標の一つで、広告の表示（掲載）回数のこと。広告枠が設定されたWebページを訪問者が閲覧し、広告が1回表示されることを1インプレッションという。“imp”あるいは“imps”（複数形）と略記されることもある。広告を掲載する媒体（メディア）側にとっては、Webページの閲覧回数（ページビュー）にページあたりの広告枠の数を乗じたものがサイトの合計インプレッションとなる。各ページに平均で3つの掲載枠があるサイトならばページビューの3倍が広告インプレッション数となる。
広告主側にとっては、掲載を依頼した広告クリエイティブ（バナー画像など）が閲覧者のWebブラウザに表示された回数がインプレッションとなる。通常、複数の広告枠があるページでも同じ広告が同時に掲載されることは稀であるため、掲載ページの合計ページビューとインプレッションはほぼ等しくなることが多い。
インプレッション1回あたりで課金する方式を「インプレッション課金型広告」と呼び、慣例として単価を1000インプレッションあたりの価格（CPM：Cost Per Mille）で表す。「CPM課金型」「PV課金型」とも呼ばれる。(http://e-words.jp/w/%E3%82%A4%E3%83%B3%E3%83%97%E3%83%AC%E3%83%83%E3%82%B7%E3%83%A7%E3%83%B3.html)
11 訳注：ペタバイト 1000兆バイト。よく耳にする「テラバイト」は1兆バイト。
12 訳注：ユニーク 日常語としての意味とは異なり、データや識別名などについて、重複のない（重複を取り除いた）、一意の、唯一の、固有の、といった意味を表すのが一般的である。例えば、「ユニークなユーザー名を設定する必要があります」というのは風変わりで個性的な名前を付けろという意味ではなく、他のユーザーと重ならない固有の名前でなければならないという意味である。(http://e-words.jp/w/%E3%83%A6%E3%83%8B%E3%83%BC%E3%82%AF.html))))

「af64a09c2」として知られる人物（住んでいる場所、読んだもの、購入したもの）について十分なプロファイルが構築できるようになりますから、普通に言われる意味での名前は必要ありません。トラッカーは、Cookieやモバイル広告IDなどの人工識別子を使用して、ターゲットを絞ったメッセージの送受信によってユーザーに接近することができます。また、実名に関連付けられていないデータも同様に機密性がないとはいえないことになります。個人情報の「匿名」プロファイルは、ほぼ常に実在の人物にリンクすることが可能なのです。

Cookie[13]などの一部の種類の識別子は、私たちが用いる技術のなかに組み込まれた機能です。その他に、ブラウザ・フィンガープリント[14]のようなものは、こうした技術が機能する仕組みから登場してきたものです。このセクションでは、Webおよびモバイルアプリのトラッカーがデータを識別および属性化する方法について説明します。

このセクションでは、サードパーティのトラッカーが使用できる識別子の代表的なサンプルについて説明します。 網羅的ではありません。ユーザーを特定しようとするトラッカーには、わたしたちが理解している以上に多くの方法があり、技術の進化とともに新しい識別子が出現しています。以下の表は、各タイプの識別子がどのようにユニークで永続的に利用可能であるかについての簡単な概要を示しています。

ウェブ識別子	ユニーク	永続的	利用可能
クッキー	はい	ユーザが削除するまで	追跡防止のないブラウザの場合
IPアドレス	はい	同じネットワーク上で、数週間または数か月持続する場合がある	常時可能
TLS状態	はい	1週間まで	ほとんどのブラウザで可能
ローカルストレージスーパーCookie	はい	ユーザーが削除するまで	サードパーティのIFrame[15]のみ。トラッカーブロッカーによってブロック可能

13 訳注：cookie クッキー Webサイトの提供者が、Webブラウザを通じて訪問者のコンピュータに一時的にデータを書き込んで保存させる仕組み。Cookieには Webサイト（Webサーバ）側が指定したデータを保存しておくことができ、利用者の識別や属性に関する情報や、最後にサイトを訪れた日時などを記録しておくことが多い。ネットサービスなどのサイトで利用者の IDなどが保存されると、次にアクセスしたときに自動的に利用者の識別が行われ、前回の続きのようにサービスを受けることができる。(http://e-words.jp/w/Cookie.html)

14 訳注： フィンガープリント 人物や端末などの識別や同定、真正性の確認に用いられる短いデータ列などを指す。例えば、電子メールのようなメッセージをインターネットなど信頼できない経路で伝送する際に、本文をハッシュ関数で計算して得られたハッシュ値を末尾などに記載したものをフィンガープリントという。受信者は受け取ったメッセージの本文から同じようにハッシュ値を計算し、添付されたフィンガープリントに一致するか比較することで、伝送途上で攻撃者による改竄やすり替えが行われていないか確認することができる。これに公開鍵暗号を組み合わせ、送信者の本人確認なども行えるようにしたものを「デジタル署名」（電子署名/公開鍵署名）という。(http://e-words.jp/w/%E3%83%95%E3%82%A3%E3%83%B3%E3%82%AC%E3%83%BC%E3%83%97%E3%83%AA%E3%83%B3%E3%83%88.html)

15 訳注： IFrame インラインフレーム。HTMLの要素（タグ）の一つで、Webページ内に矩形の領域を設け、別の Webページなどを読み込んで表示するもの。

ウェブ識別子	ユニーク	永続的	利用可能
ブラウザ・フィンガープリント	特定のブラウザのみ	はい	ほとんど常時；通常、JavaScript アクセスが必要。トラッカーブロッカーによってブロックされることもある

電話識別子	ユニーク	永続的	利用可能
電話番号	はい	ユーザーが変更するまで	データブローカーから簡単に入手可能。特別な権限を持つアプリでのみ表示可能
IMSI および IMEI 番号	はい	はい	特別な権限を持つアプリにのみ表示可能
広告 ID	はい	ユーザーがリセットするまで	はい、すべてのアプリで可能
Mac アドレス	はい	はい	アプリに対して：特別な権限を持つ場合のみ。パッシブトラッカーに対して：OS がランダム化を実行するか、デバイスが機内モードでない限り表示可能

その他の識別子	ユニーク	永続的	利用可能
ナンバープレート	はい	はい	はい
フェイスプリント	はい	はい	はい
クレジットカード番号	はい	数か月または数年は、はい	支払い処理に関与する企業には可能

3.1. Web 上の識別子

ブラウザは、ほとんどの人が Web を操作する場合の主要な手段です。Web サイトにアクセスするたびに、そのサイトのコードが原因となって、ブラウザでは表示されないサードパーティに対して数十または数百のリクエスト[16]を行う場合があります。各リクエストに

16 訳注： リクエスト データの送信や処理を要求する操作や処理、メッセージなどのことをリクエストということが多い。“req”や“rq”などの略号で示されることもある。リクエストに対する返信、応答な

は、あなたを追跡するのに利用できる情報がいくつか含まれています。

3.1.1. リクエストの構造

あなたのブラウザーと Web サイトのサーバーとの間で送受信されるほとんどすべてのデータは、HTTP リクエストの形式です。基本的に、ブラウザは Web サーバーに特定の URL を送信してコンテンツを要求します。Web サーバーは、テキストや画像などのコンテンツで応答するか、リクエストを受信したことを簡単に確認して応答します。Cookie で応答することもできます。Cookie には、追跡のために一意の識別子を含めることができます。

あなたがアクセスする各 Web サイトは、数十または数百の異なるリクエストを開始します。ブラウザのアドレスバーに表示される URL は最初のリクエストのアドレスですが、他の何百ものリクエストはバックグラウンドで行われます。これらのリクエストは、画像、コード、スタイル[17]のロード、または単にデータの共有のために利用しうるものです。

図 3.1.1: URL の一部。 ドメインは、リクエストの送信先をコンピューターに指示しますが、パスとパラメーターは、受信サーバーが必要に応じて解釈しうるような情報を伝達します。

URL 自体には、いくつかの異なる情報が含まれています。最初の部分は「nytimes.com[18]」のようなドメインです。これにより、ブラウザと接続するサーバーが指示されます。次は path で、ドメインの末尾にある「/section/world.html」のような文字列です。nytimes.com のサーバーは、パスの解釈方法を選択しますが、通常、配信するコンテンツの一部 (この場合はワールドニュースセクション) を指定します。URL の最後には「?key1 = value1 & key2 = value2」の形式のパラメーター[19]があります。通常、パラメーターには、ユーザーが行ったクエリ、ページに関するコンテキスト、追跡識別子など、リクエストに関する追加情報が含まれます。

どのことは「レスポンス」(response) あるいは「リプライ」(reply) と呼ばれる。(http://e-words.jp/w/%E3%83%AA%E3%82%AF%E3%82%A8%E3%82%B9%E3%83%88.html)

17 訳注： スタイル Web のデザインや構成などのこと。

18 訳注：これはニューヨークタイムスのウエッブの URL。

19 訳注： パラメーター ソフトウェアやシステムの挙動に影響を与える、外部から投入されるデータなどのこと (http://e-words.jp/w/%E3%83%91%E3%83%A9%E3%83%A1%E3%83%BC%E3%82%BF.html)

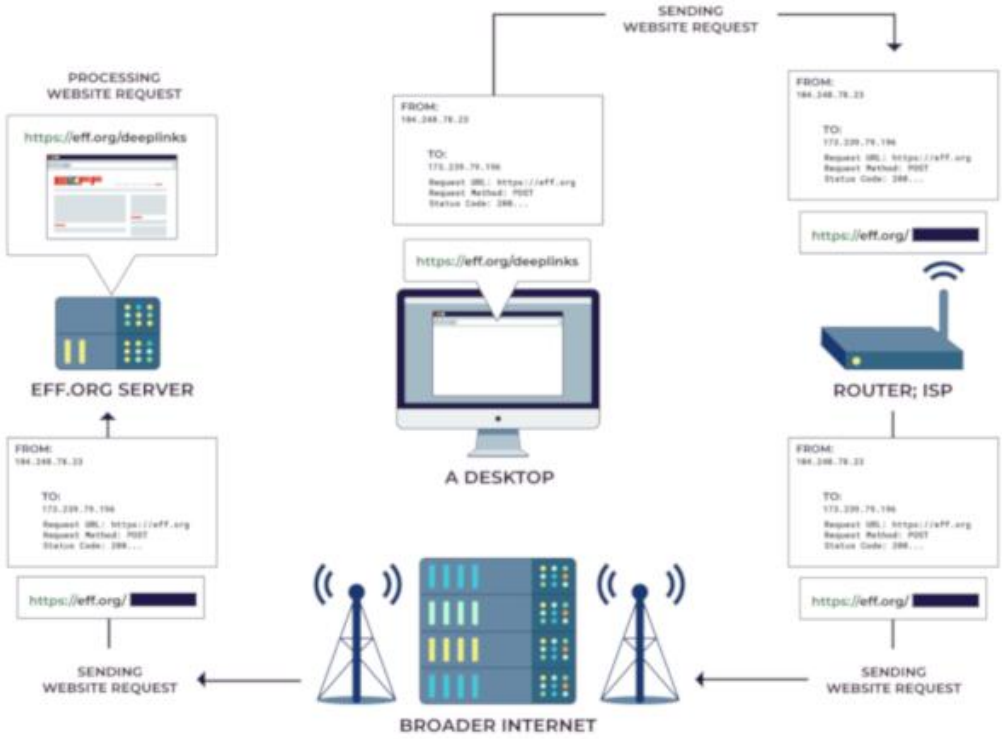

図 3.1.2: リクエストのパス。 マシンを離れた後、要求はルーターによって ISP にリダイレクトされ、ISP が「インターネット」の一連の中間ルーティングステーションを介して送信します。最後に、ドメインによって指定されたサーバーに到着します。または場合) 応答する。

サーバーに送信されるのは URL だけではありません。 HTTP ヘッダーもあります。これには、デバイスの言語やセキュリティ設定、「参照」URL、Cookie などのリクエストに関する追加情報が含まれています。たとえば、User-Agent ヘッダーは、ブラウザのタイプ、バージョン、およびオペレーティングシステムを識別します。IP アドレスや共有暗号化状態など、接続に関する下位レベルの情報もあります。一部のリクエストには、POST データの形式でさらに構成可能な情報が含まれます。POST リクエストは、Web サイトが大きすぎて扱いにくくて URL に収まらないデータの塊を共有する方法です。ほぼ何でも含めることができます。

URL や POST データなどのこの情報の一部は、個々のリクエストごとに特別に調整されています。IP アドレスや Cookie などの他の部分は、マシンによって自動的に送信されます。ほとんどすべてを追跡に利用できます。

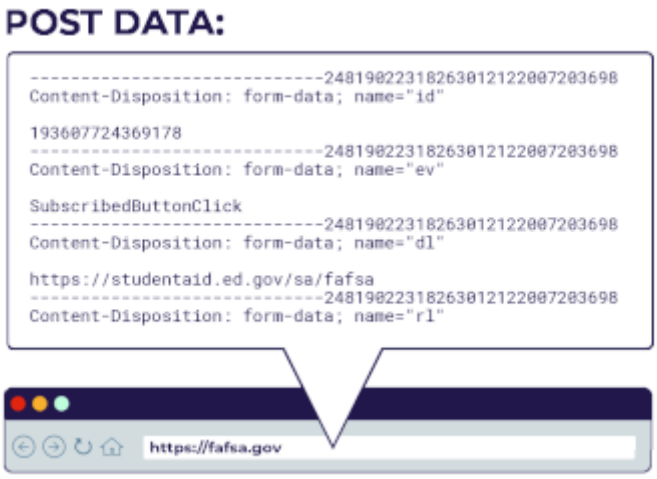

図 3.1.3: (アニメーションの解説)バックグラウンドリクエストに含まれるデータ。この画像では、ユーザーは fafsa.gov にナビゲートしていますが、ページはバックグラウンドで facebook.com へのサードパーティのリクエストを引き起こします。受信サーバーに送信される情報は URL だけではありません。HTTP ヘッダーにはユーザーエージェント文字列や Cookie などの情報が含まれ、POST データにはサーバーが必要とするものをすべて含めることができます。

上のアニメーションには、Firefox の通常バージョンから直接収集したデータが含まれています。自分でチェックアウトしたい場合は、可能です。すべての主要なブラウザには「インスペクター」または「開発者」モードがあり、ユーザーは特定のタブからのすべてのリクエストを含め、舞台裏で何が起こっているのかを見ることができます。Chrome および Firefox では、Crtl + Shift + I (または Mac では⌘+ Shift + I) を使用してこのインターフェイスにアクセスできます。「ネットワーク」タブには、特定のページによって行われたすべてのリクエストのログがあり、各ページをクリックして、そのページの行き先と含まれている情報を確認できます。

3.1.2. 自動的に共有される識別子

一部の識別可能な情報は、各リクエストとともに自動的に共有されます。 これは、必要に応じて (インターネットを駆動する基盤となるプロトコルに必要な IP アドレスと同様に)、あるいはデザインによって、Cookie と同様に、必要です。 ここで説明されている情報を収集するために、トラッカーはリクエストやリクエストをトリガーする以外に何もする必要はありません。

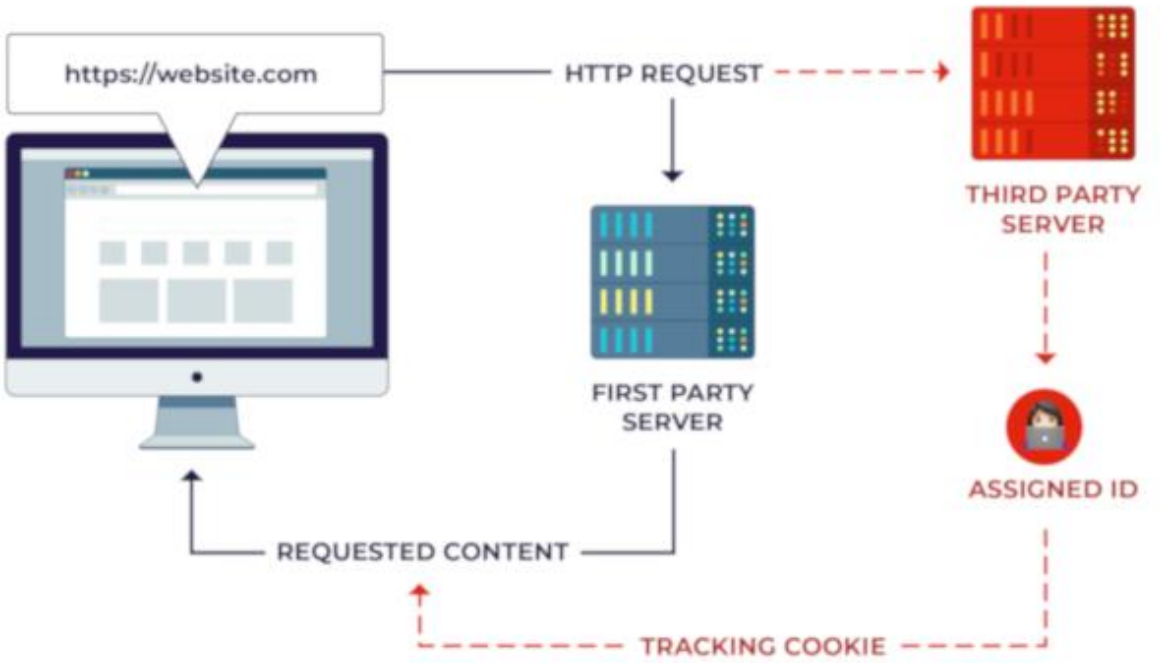

図 3.1.4: URL を入力するか、リンクをクリックして Web サイトにアクセスするたびに、コンピューターはその Web サイトのサーバー（「ファーストパーティ」）にリクエストを行います。また、他のサーバーに数十または数百のリクエストを送信する場合があり、その多くがあなたを追跡できる場合があります。

3.1.3. クッキー

サードパーティトラッキングの最も一般的なツールは HTTP Cookie です。Cookie は、特定のドメインに関連付けられた、ブラウザーに保存される小さなテキストです。Cookie は、Web サイトの所有者が、アクセスしてきたユーザーが以前に自分のサイトにアクセスしたことがあるかどうかを判断できるようにするために考案されたもので、行動追跡に最適なものです。仕組みは次のとおりです。

ブラウザーがドメイン（www.facebook.com など）に初めて要求を行うとき、サーバーはこれへの応答に際して Set-Cookie ヘッダーを添付することができます。これによって、Web サイトが望む値[20]を保存するようあなたのブラウザに指示します。たとえば、ċ$_{user}$："100026095248544 "`(これは本論文の作成者のブラウザから取得した実際の Facebook Cookie）です。あなたのブラウザは、その後 www.facebook.com にリクエストを行うたびに、以前に設定された Cookie も送信されます。このようにして、Facebook はリクエストを受け取るたびに、どのユーザーまたはデバイスからのものかを認識します。

20 訳注：値 コンピュータプログラム中で処理の対象となる単一のデータ。何らかのデータ型を持ち、一定の書式や制約に従って記述・処理される。(http://e-words.jp/w/%E5%80%A4.html)

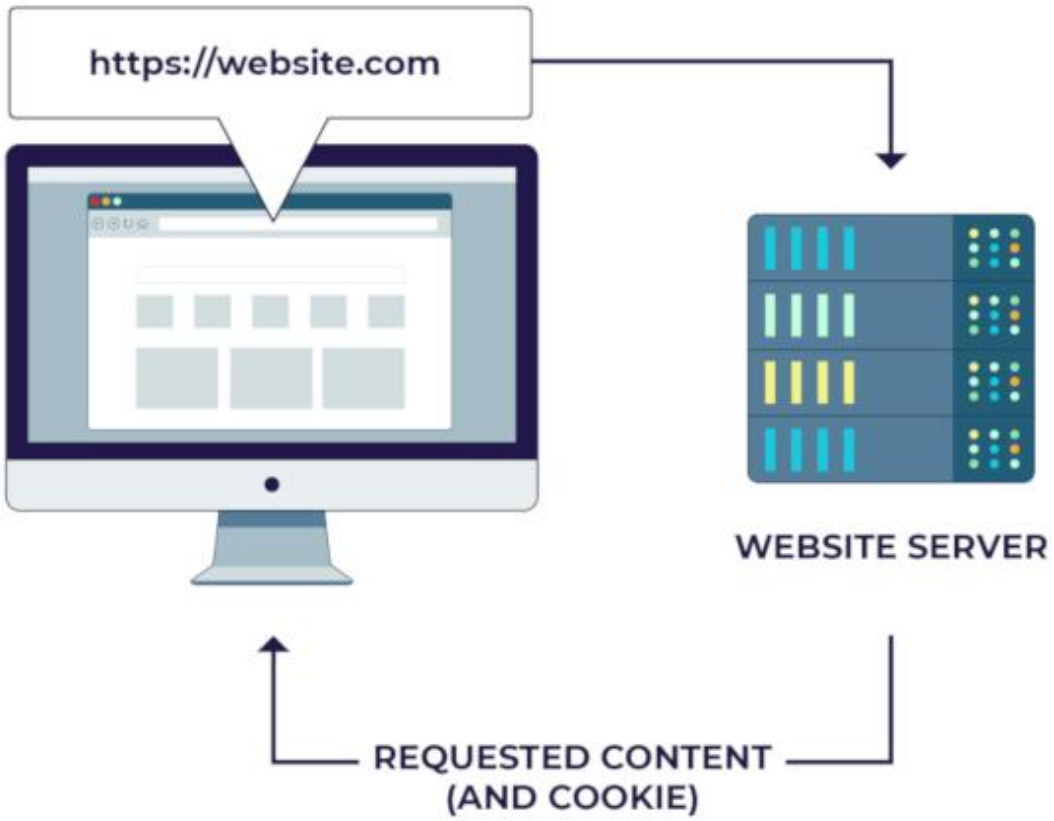

図 3.1.5: ブラウザーが新しいサーバーに初めて要求を行い、このサーバーはブラウザーにトラッキング Cookie を保存する「Set-Cookie」ヘッダーで応答することができるようになる。

すべての Cookie がトラッカーなわけではありません。ウェブサイトにアクセスするたびにログインする必要がないのは、Cookie に起因するものです。ショッピングの途中でウェブサイトを離れてもカートが空にならない理由もこのためです。Cookie は、あなたのブラウザからアクセスしている Web サイトと情報を共有するための手段にすぎません。ただし、トラッキング情報を伝送できるように設計されており、サードパーティによるあなたのトラッキングという利用法が最も悪名高いものです。

幸いなことに、ユーザーはブラウザーが Cookie を処理する方法を十分に制御できます。すべての主要なブラウザには、サードパーティの Cookie を無効にするオプション設定があります（通常、デフォルトではオフになっています）。さらに、 Safari と Firefox は、トラッカーと見なされるドメインのサードパーティ Cookie へのアクセス制限を最近開始しました。トラッカーとそれらをブロックする方法との間のこの「猫と鼠のゲーム」の結果として、サードパーティのトラッカーは、ユーザーを識別するためにクッキーだけに依存することをやめて、他の識別子に依存するように進化しています。

Cookie は常に一意であり、通常、ユーザーが手動でクリアするまで保持されます。トラッカーは、古いバージョンの Chrome では常に Cookie を使用できますが、Safari および Firefox では、多くのトラッカーはサードパーティの Cookie を使用できなくなりました。ユーザーは、ブラウザ拡張機能を使用して自分でいつでも Cookie をブロックできます。

3.1.4. IP アドレス

インターネットを介して行う各リクエストには、デバイスに固有の一時的な識別子である

IP アドレスが含まれています。 一意ですが、必ずしも永続的ではありません。新しいネットワークに移動するたびに（たとえば、自宅から職場、コーヒーショップに移動するたびに）IP アドレスが変わります。IP アドレスが機能する仕組みによって、同じネットワークに接続したままであっても、IP アドレスは変わりうるのです。

広く使用されている IP アドレスには、IPv4 と IPv6 の 2 種類があります。IPv4 は、Web よりも 10 年前の技術です。わずか数百の機関で使用されるインターネット向けに設計されており、現在世界で接続されている 220 億以上のデバイスに対応できる IPV4 アドレスは世界で約 40 億個しかありません。それでも、インターネットトラフィックの 70％以上が IPv4 を使用しています。

その結果、消費者用のデバイスで使用される IPv4 アドレスは常に再割り当てをされています。デバイスがインターネットに接続すると、そのインターネットサービスプロバイダー（ISP）は、IPv4 アドレスで「リース」します。 これにより、デバイスは単一のアドレスを数時間または数日間使用できます。リースが有効になると、ISP はリースを延長するか、新しい IP を付与するかを決定できます。デバイスが長期間同じネットワーク上にある場合、その IP は数時間ごとに変更される場合があります。または、数か月間変更されない場合もあります。

IPv6 アドレスには IPv4 のような希少性の問題はありません。変更する必要はありませんが、技術標準のプライバシーを保護する拡張機能によって、ほとんどのデバイスは数時間または数日ごとに新しいランダムな IPv6 アドレスを生成します。つまり、IPv6 アドレスは短期的な追跡や他の識別子のリンクに使用できますが、スタンドアロンの長期的な識別子としては使用できません。

IP アドレスはそれ自体では完全な識別子ではありませんが、十分なデータがあれば、トラッカーはそれらを利用して、デバイス間のマッピング関係など、ユーザーの長期プロファイルを作成できます。信頼できる VPN または Tor ブラウザを使用して、サードパーティのトラッカーから IP アドレスを隠すことができます。

IP アドレスは常に一意であり、ユーザーが VPN または Tor[21]を介して接続しない限り、トラッカーが常に利用できます。IPv4 アドレスも IPv6 アドレスも数日間以上持続することは保証されていませんが、IPv4 アドレスは数か月間持続する場合があります。

3.1.5. TLS 状態

現在、Web 上のほとんどのトラフィックは、トランスポートレイヤーセキュリティ（TLS）[22]を使用して暗号化されています。「https：//」で始まる URL に接続するときはいつでも、TLS を使用して接続しています。これはとてもいいことです。TLS と HTTPS が提供する暗号化された接続は、ISP、ハッカー、政府が Web トラフィックをスパイするのを防ぎ、データが宛先に向かう途中で傍受または変更されないようにします。

ただし、トラッカーがユーザーを識別するための新しい方法にも道を開きます。TLS セッ

21 訳注：Tor 第 4 章参照。日本語の解説として「Tor プロジェクト解説」も参照。
22 訳注： TSL インターネットなどの TCP/IP ネットワークでデータを暗号化して送受信するプロトコル（通信手順）の一つ。データを送受信する一対の機器間で通信を暗号化し、中継装置などネットワーク上の他の機器による成りすましやデータの盗み見、改竄などを防ぐことができる。(http://e-words.jp/w/TLS.html)

ション ID[23]とセッションチケット[24]は、暗号化された接続の高速化に役立つ暗号化識別子です。HTTPS 経由でサーバーに接続すると、ブラウザはサーバーと新しい TLS セッションを開始します。

セッションのセットアップには、いくつかのかなり負担の大きな暗号化の作業が含まれるため、サーバーは必要以上に頻繁にこうした作業を行うのを好みません。再接続するたびにサーバーとブラウザの間で完全な暗号化「ハンドシェイク」[25]を実行する代わりに、サーバーは共有された暗号化の状態の一部をエンコードするセッションチケットをブラウザに送信することができます。次回同じサーバーに接続するときに、ブラウザーはセッションチケットを送信し、ハンドシェイクを省略できるようにします。唯一の問題点として、セッションチケットがトラッカーによって一意の識別子として悪用される可能性があります。

TLS セッショントラッキングは最近、ある学術論文で一般の注目を集めましたが、その利用がどのくらい普及しているかははっきりしていません。

IP アドレスと同様に、セッションチケットは常に一意です。Tor のように、ユーザーのブラウザが拒否するように設定されていない限り、それらは利用可能です。通常、サーバー管理者はセッションチケットを最長 1 週間維持するように構成可能ですが、ブラウザーはしばらくしてからそれらをリセットします。

3.1.6. トラッカーによって作成された識別子

Web ベースのトラッカーは、IP アドレス (信頼性が低く、永続的ではない)、Cookie (ユーザーがクリアまたはブロックできる)、または TLS 状態 (数時間または数日以内に期限切れになる) 以外の識別子を使用しようとすることがあります。このために、トラッカーは、更に工夫する必要があります。JavaScript[26]を利用して、 ローカルストレージにデータを保存およびロードしたり、ブラウザーのフィンガープリントを実行したりすることができます。

1. ローカルストレージの「Cookie」と IFrame

   ローカルストレージ[27]は、ウェブサイトがブラウザにデータを長期間保存する方法です。ローカルストレージは、Web ベースのテキストエディターで設定を保存したり、オンラインゲームで進行状況を保存したりするのに役立ちます。Cookie と同

23 訳注：セッション ID Web アプリケーションなどで、通信中の利用者を識別して行動を捕捉し、利用者ごとに一貫したサービスを提供するために付与される固有の識別情報。そのような仕組みをセッション管理という。コンピュータ上で直接実行されるソフトウェアの場合は OS などがセッションを管理するためソフトウェア側での制御は不要なことが多いが、Web では HTTP 自体にセッションの識別・管理のための仕組みが存在しないため、複数の利用者がアクセスしてきた場合にこれを識別する仕組みが必要となる。
そのような場合に発行されるのがセッション ID で、Web サーバと Web ブラウザで情報を共有する仕組みである Cookie などを応用し、サーバが初回アクセス時にセッション ID を発行してクライアントが保存する。以降は通信のたびにクライアント側からセッション ID を申告することで、同時にアクセス中の利用者の識別・同定を行う。
利用者自体を継続的に識別するユーザー名などの識別子とは異なり、機械的に生成されて一時的に利用されるもので、一連の通信が終了すると破棄される。同じ利用者が次に通信を開始すると新しいセッション ID が与えられる。(http://e-words.jp/w/%E3%82%BB%E3%83%83%E3%82%B7%E3%83%A7%E3%83%B3ID.html)

24 訳注：セッションチケット 「Server から送られて来た暗号化された Session 情報は Client は復号できないため単なるバイトデータに過ぎません。“暗号化された Session 情報”は Session Ticket と呼ばれます。この仕組みが有効に機能する場合、Session の再開は Server が Session 情報を保存しなくても実現する事ができます。」(「細かすぎて伝わらない SSL/TLS」https://techblog.yahoo.co.jp/infrastructure/ssl-session-resumption/) この論文にはかなり詳細な解説が含まれている。

25 訳注：ハンドシェイク 二台の装置が通信する際に、相手方の応答確認や設定値の交換など、本来的な信号やデータの伝達以外に行われる通信のこと。(http://e-words.jp/w/%E3%83%8F%E3%83%B3%E3%83%89%E3%82%B7%E3%82%A7%E3%82%A4%E3%82%AF.html)

26 訳注： Javascript 主に Web ページに組み込まれたプログラムを Web ブラウザ上で実行するために用いられるプログラミング言語の一つ。(http://e-words.jp/w/JavaScript.html)

27 訳注：ローカルストレージ 現在操作している手元のコンピュータに内蔵あるいは直に繋がれたハードディスクなどの外部記憶装置 (ストレージ) のこと。また、その中に設けられた特定のソフトウェアのためのデータ保存領域。(http://e-words.jp/w/%E3%83%AD%E3%83%BC%E3%82%AB%E3%83%AB%E3%82%B9%E3%83%88%E3%83%AC%E3%83%BC%E3%82%B8.html)

様に、ローカルストレージを使用すると、サードパーティのトラッカーがブラウザで一意の識別子を作成して保存できます。

Cookie と同様に、ローカルストレージのデータは特定のドメインに関連付けられています。つまり、example.com がブラウザーに値を設定すると、example.com Web ページと example.com の IFrame だけがそこにアクセスできます。 IFrame は、Web ページ内の小さな Web ページのようなものです。IFrame 内では、サードパーティドメインはファーストパーティドメインが実行できるのとほぼすべて同じことを実行できます。たとえば、埋め込み YouTube ビデオは IFrame を使用して構築されています。YouTube 以外のサイトで YouTube ビデオを見るたびに、小さなページ内ページでこれが実行されているのです。ほとんどの場合、ブラウザは YouTube IFrame を本物の Web ページのように扱い、YouTube のローカルストレージの読み取りと書き込みを許可します。確かに、YouTube はそのストレージを使用して一意の「デバイス識別子」を保存し、ビデオが埋め込まれたページでユーザーを追跡します。

ローカルストレージの「Cookie」は一意であり、ユーザーがブラウザのストレージを手動でクリアするまで保持されます。これらは、サードパーティの IFrame 内で JavaScript コードを実行できるトラッカーでのみ利用できます。すべての Cookie ブロック手段がローカルストレージ Cookie を考慮するとは限らないため、ローカルストレージ Cookie は、通常の Cookie アクセスがブロックされるトラッカーによって利用される可能性があります。

2. 指紋(フィンガープリンティング)

ブラウザのフィンガープリントは、Web ベースの追跡のなかでも最も複雑で潜行性のある形式のひとつです。ブラウザフィンガープリントは、単独または組み合わせて、個々のデバイス上の個々のブラウザを一意に識別する 1 つ以上の属性から構成されています。通常、フィンガープリントに入るデータは、ブラウザーが Web とやり取りする方法の一部にすぎないため、ブラウザーが露出せざるをえないものです。これらには、ブラウザがサイトを訪問するたびに行われるリクエストと一緒に送信される情報や、ページで JavaScript を実行することで発見できる属性が含まれます。たとえば、あなたのモニターの画面の解像度、インストールしたソフトウェアの特定のバージョン、タイムゾーンが含まれます。あなたのブラウザーのアクセス先の Web サイトに露出させる情報は、ブラウザーのフィンガープリントを作成するのに利用することができます。EFF の Panopticlick プロジェクトを利用すると、自分のブラウザのフィンガープリントがどんなものかなんとなくわかると思います。。

フィンガープリンティングの信頼性は活発に研究されているテーマであり、進化し続ける Web テクノロジーをもとに評価する必要があります。ただし、新しい手法によって一意の識別の可能性が高まり、フィンガープリントを使用するサイトの数も増えているのは明らかです。最近のレポートでは、アメリカ人がアクセスした上位 500 サイトのうち少なくとも 3 分の 1 が何らかの形式のブラウザフィンガープリンティングを採用していることがわかっています。サイトでのフィンガープリントの普及率は、Web サイトのカテゴリによっても大きく異なります。

研究者は、 キャンバス・フィンガープリント技術がブラウザの識別に特に効果的であることを発見しました。HTML Canvas は HTML5 の機能であり、Web ページが Web ページ内に複雑なグラフィックを表示できるようにします。ゲーム、アートプロジェクト、および Web の美しさを重視したいくつかのサイトで使用されています。非常に複雑でパフォーマンス重視のため、デバイスごとに動作が少し異なります。キャンバスのフィンガープリントはこれを利用します。

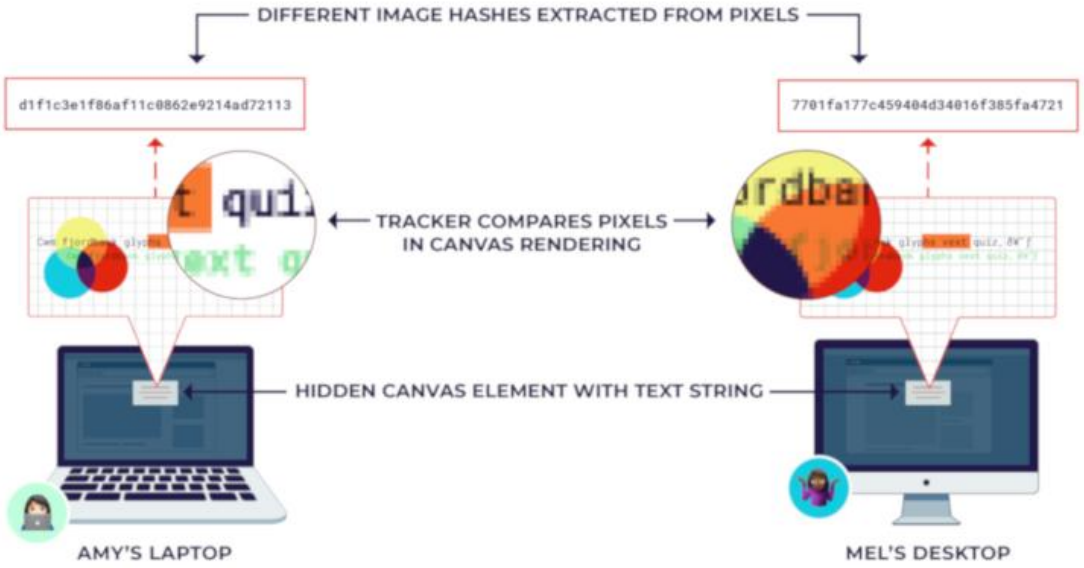

図 3.1.6: キャンバスフィンガープリント。 トラッカーは、さまざまなフォントで図形、グラフィック、およびテキストをレンダリングし、描画されるピクセルの「ハッシュ」を計算します。ハッシュは、ハードウェア、ファームウェア、またはソフトウェアのわずかな違いのあるデバイスでも異なるものになります。

トラッカーは、ユーザーには見えない「キャンバス」要素を作成し、JavaScript を使用して複雑なシェイプ[28]またはテキスト文字列をレンダリングし、キャンバス上の各ピクセルがどのようにレンダリング[29]するのかに関するデータを正確に抽出することができます。コンピューターにインストールされているオペレーティングシステム、ブラウザーバージョン、グラフィックカード、ファームウェアバージョン、グラフィックドライバーバージョン、およびフォントはすべて最終結果に影響します。

フィンガープリンティングの目的からみて、個々の特性は単独で測定されることはほとんどありません。トラッカーは、最も効果的に、複数の特性を組み合わせて、残された情報の断片をまとまりのある全体につなぎ合わせて、ブラウザーを識別します。キャンバス・フィンガープリントのようなひとつの特性だけではブラウザを一意に識別できない場合でも、通常は、他の特性（言語、タイムゾーン、ブラウザ設定）と組み合わせて識別することがきます。ですから、情報の単純な値の組み合わせの利用は、あなたの想像以上にはるかに効果的なのです。

フィンガープリントはしばしば一意ですが、常にそうだというわけではありません。Tor や Safari などの一部のブラウザーは、ユーザーが同じように見えるように特別に設計されており、ブラウザーフィンガープリントの有効性が削除または制限されています。ブラウザのフィンガープリントは、ユーザーが同じハードウェアとソフトウェアを持っている限り持続する傾向があります。フィンガープリントを「リセット」するように調整できる設定はありません。また、フィンガープリントは通常、ブラウザで JavaScript を実行できるサードパーティによって利用モできます。

28 訳注： シェイプ 画像編集ソフトや文書作成ソフトなどで、点や線、面を組み合わせて作成された図形のこと(http://e-words.jp/w/%E3%82%B7%E3%82%A7%E3%82%A4%E3%83%97.html)

29 訳注：レンダリング 画像や画面の内容を指示するデータの集まり（数値や数式のパラメータ、描画ルールを記述したものなど）をコンピュータプログラムで処理して、具体的な画素の集合を得ること。3 次元グラフィックスで数値データとして与えられた物体や図形に関する情報を計算によって画像化することを指すこと。(http://e-words.jp/w/%E3%83%AC%E3%83%B3%E3%83%80%E3%83%AA%E3%83%B3%E3%82%B0.html)

3.1.7. モバイルデバイス上の識別子

スマートフォン、タブレット、および電子書籍リーダーには、通常、デスクトップブラウザーと同じように機能する Web ブラウザーをもっています。つまり、これらのタイプの接続デバイスは、上記のセクションで説明したすべての種類のトラッキングの影響を受けやすいのです。

ただし、モバイルデバイスにはふたつの大きな違いがあります。まず、ユーザーは通常、デバイスの機能を最大限に活用するために Apple、Google、または Amazon アカウントでサインインする必要があります。これにより、デバイス識別子がアカウント ID にリンクされ、これらの強力な企業はユーザーの行動をプロファイリングしやすくなります。たとえば、自宅や職場の住所を Google マップに保存するには、Google の「ウェブとアプリのアクティビティ」をオンにする必要があります。これにより、場所、検索履歴、アプリのアクティビティがターゲット広告によって利用できるようになりす。

第二に、同様に重要なことですが、ほとんどの人はほとんどの時間をモバイルデバイスでブラウザー以外のアプリを使って過ごします。アプリ内のトラッカーは、Web ベースのトラッカーと同じ方法では Cookie にアクセスできません。ただし、モバイルオペレーティングシステムの機能を活用することで、アプリトラッカーは一意の識別子にアクセスして、アクティビティをデバイスに関連付けることができます。さらに、携帯電話、特に Android および iOS オペレーティングシステムを実行している携帯電話は、追跡に利用できる一意の識別子の組み合わせへのアクセスを持ちます。

モバイルエコシステム[30]では、ほとんどの追跡はサードパーティのソフトウェア開発キット（SDK）によって行われます。SDK[31]は、アプリ開発者がアプリに含めることを選択できるコードのライブラリです。前述のように、SDK はほとんどの場合、サードパーティが活用する Web リソースと同じように機能します。SDK はサードパーティがあなたの動作、デバイス、その他の特性を学習できるようにします。サードパーティの分析サービスを利用したり、サードパーティの広告を配信したりするアプリ開発者は、Google や Facebook などからコードをダウンロードします。そして、この開発者はアプリの公開バージョンにそのコードを含めます。したがって、サードパーティのコードは、場所やカメラへのアクセスなど、アプリに付与された権限によって保護されているデータを含め、アプリが行うすべてのデータにアクセスできます。

PC では、Web では、ブラウザは「ファーストパーティ」リソース[32]と「サードパーティ」リソースの区別を実行します。これにより、ブラウザストレージへのアクセスをブロックするなど、サードパーティのコンテンツに追加の制限を加えることができます。モバイルアプリでは、この区別は存在しません。あなたは、アプリ内で実行されているすべてのサードパーティコードに同じ特権を付与することなしに、あるひとつのアプリに特権を付与することはできません。

3.1.8. 電話番号

電話番号は、最も古い一意の数値識別子のひとつであり、最も理解しやすいもののひとつです。各番号は特定のデバイスに固有のものであり、頻繁に変更されることもありません。ユーザーは、さまざまな理由でアカウント番号、電子レシート、実店舗でのロイヤルティプログラムなどのために、電話番号を共有するように推奨されます。したがって、データブローカーは頻繁に電話番号を収集して販売します。ただし、アプリ内から電話番号にア

30 訳注：エコシステム 「IT におけるエコシステム」は、デベロッパーやベンダーが平等に利益を得られる協業形式 経営や IT での使用の事例は https://ec-orange.jp/ec-media/?p=23780

31 訳注：SDK Software Development Kit。ソフトウェア開発キット。あるシステムに対応したソフトウェアを開発するために必要なプログラムや文書などをひとまとめにしたパッケージのこと。システムの開発元や販売元が希望する開発者に配布あるいは販売する。近年ではインターネットを通じてダウンロードできるよう Web サイトで公開されることが多い。(http://e-words.jp/w/SDK.html)

32 訳注： リソース ソフトウェアやハードウェアを動作させるのに必要なコンピュータの構成要素やその能力、すなわち、CPU の処理速度やメモリ容量、ストレージ容量などを指す。通信ネットワークなどを通じて入手あるいは利用できる、外部の機器やソフトウェア、その提供する機能や容量、データ、ファイルなど、何らかの役に立つ要素を総称的にリソースと呼ぶことがある。(http://e-words.jp/w/%E3%83%AA%E3%82%BD%E3%83%BC%E3%82%B9.html)

クセスするのは簡単ではありません。Android では、電話番号は特定の権限が付与されたアプリのサードパーティトラッカーにのみ利用可能です。iOS は、アプリはユーザーの電話番号にまったくアクセスできないようにしています。

電話番号は一意で永続的ですが、通常、ほとんどのアプリでサードパーティのトラッカーは利用できません。

3.1.9. ハードウェア識別子：IMSI および IMEI

モバイルネットワークに接続できるすべてのデバイスには、International Mobile Subscriber Identity（IMSI）番号と呼ばれる一意の識別子が割り当てられています。IMSI 番号は、携帯電話会社によってユーザーに割り当てられ、SIM カードに保存されます。通常のユーザーは、SIM を変更しないと IMSI を変更できません。これが、トラッキングに理想的な識別子になります。

同様に、すべてのモバイルデバイスには、ハードウェアに「焼き付けられた」IMEI（International Mobile Equipment Identity）番号があります。 SIM カードと電話番号は変更できますが、新しいデバイスを購入しないと IMEI を変更できません。

IMSI 番号は、携帯電話基地局に接続するたびに（常に)携帯電話事業者と共有されます。世界中を移動すると、電話機は近くの基地局に ping[33]を送信して、ネットワークの状態に関する情報を要求します。電話会社はこの情報を使用して、位置を（さまざまな精度で）トラッキングできます。これは、あなたが関係する電話会社によって実行されているため、サードパーティの追跡ではありませんが、多くのユーザーはこうしたことが起こっていることに気付かない場合があります。

携帯電話で実行されているソフトウェアとアプリは、IMSI および IMEI 番号にもアクセスできますが、それほど簡単ではありません。モバイルオペレーティングシステムは、ユーザが許可し後で取り消すことができる権限によってハードウェア識別子へのアクセスをロックします。たとえば、Android Q 以降、アプリはリセット不可の ID を読み取るために「READ$_{\text{PRIVILEGEDPHONESTATE}}$」[34]権限を要求する必要があります。iOS では、アプリがこれらの識別子にアクセスすることはまったくできません。このため、これ以外の識別子が、ほとんどのアプリベースのサードパーティのトラッカーにとってより魅力的なオプションになります。電話番号と同様に、IMSI および IMEI 番号は一意で永続的ですが、ほとんどのトラッカーは、アクセスするのに苦労するため、簡単には入手できません。

3.1.10. 広告 ID

広告 ID は、モバイルデバイスを一意に識別する長いランダムな文字と数字の文字列です。 広告 ID は技術的なプロトコルの一部ではありませんが、iOS および Android オペレーティングシステムに組み込まれています。

携帯電話の広告 ID は、Web の Cookie に似ています。ブラウザに保存され、Cookie などのさまざまな Web サイトのトラッカーと共有される代わりに、広告 ID は携帯電話に保存され、さまざまなアプリのトラッカーと共有されます。広告 ID は、行動ターゲティング広

33 訳注： ping インターネットなどの TCP/IP ネットワークで、ネットワーク上で特定の IP アドレスを持つ機器から応答があるかを調べるためのプログラム。http://e-words.jp/w/ping.html

34 訳注：READ$_{\text{PRIVILEGEDPHONESTATE}}$ 日本スマートフォンセキュリティ協会（JSSEC）の「Android アプリのセキュア設計・セキュアコーディングガイド」に下記の記述がある。「ndroid 10 では、プライバシー保護のため、再設定不可能なデバイス ID の取得がより制限されることになった。取得するためには、READ$_{\text{PRIVILEGEDPHONESTATE}}$ パーミッションが必要であるが、このパーミッションは通常のアプリには付与されない。この変更は、targetSdkVersion の指定に限らず、Android 10 上で動く全てのアプリに影響する。そのため、通常パーミッションが付与され動作していたアプリでも、SecurityException の発生や null の返却により想定している挙動とならない可能性がある。」（「Android 10 での再設定不可能なデバイス ID の取得制限について」 一般社団法人日本スマートフォンセキュリティ協会（JSSEC）セキュアコーディング WG 2019/9/1 (https://www.jssec.org/dl/android_securecoding/5_how_to_use_security_functions.html#android-10-%E3%81%A7%E3%81%AE%E5%86%8D%E8%A8%AD%E5%AE%9A%E4%B8%8D%E5%8F%AF%E8%83%BD%E3%81%AA%E3%83%87%E3%83%90%E3%82%A4%E3%82%B9id%E3%81%AE%E5%8F%96%E5%BE%97%E5%88%B6%E9%99%90%E3%81%AB%E3%81%A4%E3%81%84%E3%81%A6)

告主とデバイス上のアプリ間をいったりきたりするユーザーアクティビティとをリンクできるようにすることを唯一の目的として存在します。

IMSI または IMEI 番号とは異なり、広告 ID は変更が可能で、iOS では完全にオフにできます 。広告 ID は iOS と Android の両方で、デフォルトでは有効になっており、特別な権限なしですべてのアプリで使用できます。両方のプラットフォームで、ユーザーが手動でリセットしない限り、広告 ID はリセットされません。

Google と Apple は、開発者が IMEI や電話番号などの他の識別子の代わりに、行動プロファイリングに広告 ID を使用することを推奨しています。表面的には、ユーザーが手動で識別子をリセットできるので(これを選択すればの話ですが)、ユーザーはトラッキングの方法をより細かく制御できるようになります。ただし、実際には、ユーザーが広告 ID をリセットしたために問題を生じた場合であっても、トラッカーは、IP アドレスやアプリ内ストレージなどの他の識別子を使用することで、リセット後も簡単に識別できます。Android の開発者ポリシーでは、トラッカーにこのような動作を行わないように指示していますが、プラットフォームにはそれを止める技術的な保護手段がありません。2019 年 2 月、ある調査によると、Play ストアの 18,000 以上のアプリが Google のポリシーに違反していることが判明しています。

広告 ID は一意であり、デフォルトですべてのアプリで使用できます。ユーザーが手動でリセットするまで保持されます。このために、これらは不正なトラッカーにとって非常に魅力的な識別子になります。

3.2. MAC アドレス

インターネットに接続できるすべてのデバイスには、メディアアクセス制御（MAC）アドレスと呼ばれるハードウェア識別子があります。MAC アドレスは、WiFi または Bluetooth を介した 2 つのワイヤレス対応デバイス間の初期接続をセットアップするために使用されます。

MAC アドレスはあらゆる種類のデバイスで使用されますが、モバイルデバイスでは、これに関連するプライバシーリスクが高まります。インターネット経由でやり取りする Web サイトやその他のサーバーは、実際には MAC アドレスを見ることができませんが、住んでいる地域のネットワークデバイスは見ることができます。実際、MAC アドレスを確認するためにネットワークに接続する必要さえありません。 近くにいるだけでよいのです。

その仕組みは次のとおりです。近くの Bluetooth デバイスと WiFi ネットワークを見つけるために、デバイスはプローブ要求と呼ばれる短い無線信号を常に送信しています。各プローブ要求には、デバイスの一意の MAC アドレスが含まれています。エリア内に WiFi ホットスポットがある場合、プローブを受信して、あなたのデバイスの MAC によってアドレス指定された固有の「プローブ応答」を、接続方法に関する情報とともに送り返します。

ただし、エリア内の他のデバイスも、プローブ要求を確認および傍受できます。これは、企業がワイヤレス「ビーコン」[35]をセットアップして、周囲の MAC アドレスを盗み聞きして、そのデータを使用して特定のデバイスの動きを経時的に追跡できるということを意味しています。ビーコンは、多くの場合、企業、公開イベント、さらには政治キャンペーンの屋外看板に設置されています。十分多くの場所に十分なビーコンがあると、企業は店舗や街中のユーザーの動きを追跡できます。また、2 人が同じ場所にいることを識別し、その情報を使用してソーシャルグラフを作成することもできます。

35 訳注： ビーコン 赤外線や近距離無線通信の電波などを発して、周囲に機器の現在位置や識別情報などを知らせる小型の発信機のこと。(http://e-words.jp/w/%E3%83%93%E3%83%BC%E3%82%B3%E3%83%B3.html)

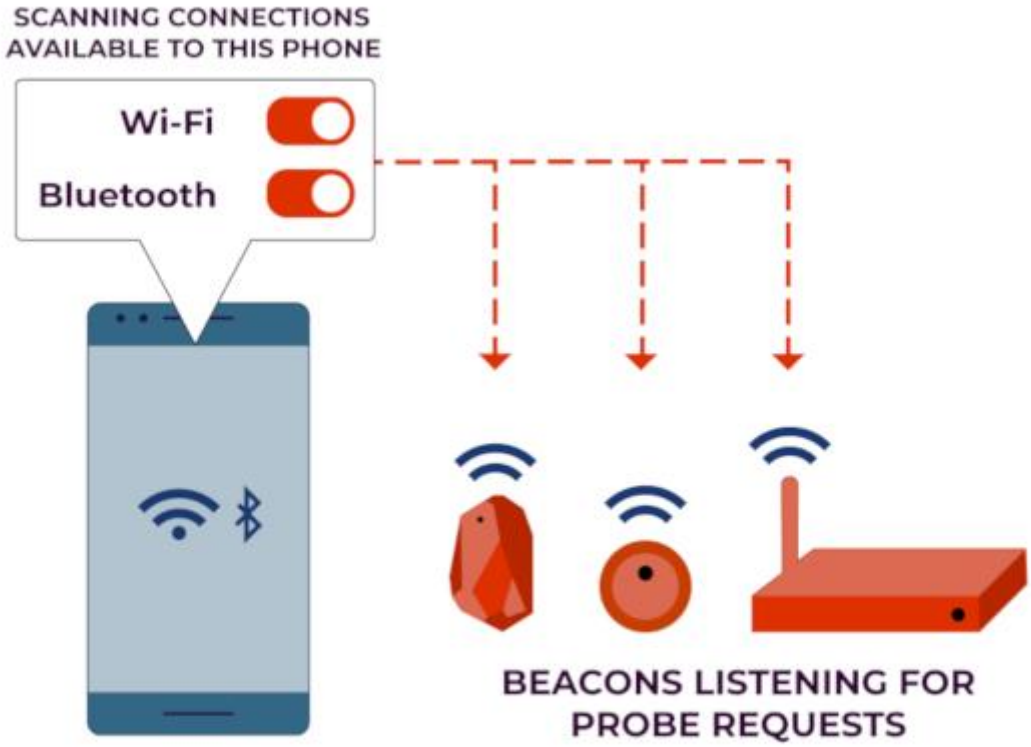

図 3.2.1: 近くの Bluetooth デバイスと WiFi ネットワークを見つけるために、デバイスはプローブ要求と呼ばれる短い無線信号を絶えず送信しています。 各プローブ要求には、デバイスの一意の MAC アドレスが含まれています。企業は、無線「ビーコン」をセットアップして、近くの MAC アドレスを盗み聞きして、そのデータを使用して、特定のデバイスの動きを経時的に追跡できます。

MAC アドレスのランダム化でこの種の追跡を阻止することができます。プローブリクエストで真のグローバルに一意の MAC アドレスを共有する代わりに、デバイスは毎回ブロードキャストする新しいランダムな「なりすまし」MAC アドレスを作成することができます。これにより、パッシブトラッカーがひとつのプローブ要求を別のプローブ要求にリンクしたり、特定のデバイスにリンクしたりすることができなくなります。幸いなことに、iOS と Android の最新バージョンには、デフォルトで MAC アドレスのランダム化が含まれています。

MAC アドレスの追跡は、ラップトップ、古い携帯電話、およびその他のデバイスのリスクとして残っていますが、業界はプライバシー保護の規範を強化する傾向にあります。

ハードウェア MAC アドレスはグローバルに一意です。また、永続的であり、デバイスの寿命の間変化しません。アプリ内のトラッカーでは簡単に利用できませんが、ワイヤレスビーコンを使用するパッシブトラッカーでは利用が可能です。ただし、多くのデバイスはデフォルトで MAC アドレスを難読化するため、パッシブトラッキングの信頼性の低い識別子になりつつあります。

3.3. 現実の識別子

多くの電子デバイス識別子は、ユーザーがリセット、難読化、またはオフにすることができます。しかし、現実世界の識別子は別の話です。運転中（および駐車中）に自動車のナ

ンバープレートを覆うことは違法であり、顔や指紋などの生体認証識別子を変更することはほぼ不可能です。

3.3.1. ナンバープレート

米国のすべての車には、実世界のアイデンティティに関連付けられたナンバープレートが法的に必要です。トラッキング識別子に関する限り、ナンバープレート番号は、それによって取得されるデータとほぼ同じものを意味しています。つまり、ナンバープレートは見つけやすく、難読化することは違法です。また、簡単に変更することはできず、どこに行ってもほとんどのすべての人についてまわります。

自動ナンバープレート読取機（ALPR）は、通過する車のナンバープレート番号を自動的に識別して記録できる専用カメラです。ALPR は、混雑した交差点やモールの駐車場などの固定地点、またはパトカーその他の車両に設置できます。民間企業は ALPR を運営し、これを利用して膨大な旅行者の位置データを収集し、このデータを他の企業（および警察）に販売しています。

残念ながら、ALPR による追跡は、運転者にとって本質的に避けられません。ナンバープレートを隠したり変更したりすることは違法です。また、ほとんどの ALPR は公共スペースで動作するため、デバイス自体を回避することは極めて困難です。

ナンバープレートは一意で、車両を見ることができる人なら誰でも見ることができ、極めて永続的です。法執行機関とサードパーティのトラッカーの両方にとって、車両とそのドライバーに関するデータを収集するための理想的な識別子です。

3.3.2. 顔生体認証

顔は、サードパーティのトラッカーにとって非常に魅力的な一意の識別子の別のクラスです。 顔はユニークで、変更することには非常に不向きです。幸いなことに、あなたの顔を一般大衆から隠すことは違法ではありませんが、ほとんどの人にとってそうすることは非現実的です。

どの顔も一意で、利用可能で、永続的です。ただし、現在の顔認識ソフトウェアは、ある顔を別の顔と混同する場合があります。さらに、 研究によれば、アルゴリズムは、有色人種、女性、高齢者を識別するときに、この種のエラーを起こしやすいことが示されています。

顔認識はすでに広く開発されていますが、その影響の広がりを感じ始めています。将来的には、顔認識カメラが店頭や街角に置かれ、コンピューター支援の眼鏡にドッキングされる可能性があります。強力なプライバシー規制がなければ、平均的な人々は、顔認識を介した一般的な追跡とプロファイリングに対抗する方法を事実上持つことができなくなるでしょう。

3.3.3. クレジット/デビットカード

クレジットカード番号は、別の優れた長期的な識別子です。サイクルアウトできますが、ほとんどの人は、Cookie をクリアするほど頻繁にクレジットカード番号を変更しません。さらに、クレジットカード番号は本名に直接結び付けられており、取引の一環としてクレジットカード番号を受け取った人は誰でもあなたの正式名を取得します。

ほとんどの人が理解できないのは、各クレジットカード取引に関係する隠れたサードパーティの規模です。地元の店でちょっとした小物を購入する場合、この店はおそらくカード処理サービスを提供する支払い処理業者と契約しています。また、取引はあなたの銀行だけでなくカード発行会社の銀行によっても確認される必要があります。支払いプロセッサは、トランザクションを検証するために他の会社と契約する場合があり、これらの会社は購入に関する情報をすべて受け取る場合があります。銀行やその他の金融機関は、データセキュリティの基準を義務付け、ユーザーデータの共有方法を開示することを義務付けら

れ、共有をオプトアウト[36]する権利をユーザーに付与するGramm-Leach-Bliley Actによって規制されています。 ただし、支払い処理業者やデータ集約業者などの他の金融テクノロジー企業は、規制が大幅に緩和されています。

3.4. 長期にわたる識別子のリンク

多くの場合、トラッカーはユーザーとの安定したリンクとして機能する単一の識別子に頼ることはできません。IP アドレスの変更、Cookie のクリア、広告 ID のリセットが可能だからです。また、より知識のあるユーザーは、プリペイドの携帯電話番号と電子メールアドレスを使用して、ID の一部を分離しようとします。こうしたことが起きた場合でも、トラッカーはあきらめることはなく、新しいユーザープロファイルを最初から作成したりはしません。その代わりに、通常、複数の識別子を組み合わせて統合プロファイルを作成します。このようにして、ある識別子または別の識別子が変更されたときにユーザーの追跡を失う可能性が低くなり、古い識別子を新しい識別子にリンクすることができるのです。

ユーザーを識別する方法は非常に多くあるため、このことがトラッカーには利点になります。ユーザーが Cookie をクリアしても IP アドレスが変わらない場合、古い Cookie を新しい Cookie にリンクするのは簡単です。あるネットワークから別のネットワークに移動し、同じブラウザーを使用する場合、ブラウザーの指紋は古いセッションを新しいセッションにリンクできます。サードパーティの Cookie をブロックし、Safari のようなフィンガープリントを採取することが難しいブラウザを使用する場合、トラッカーは、TLS セッションデータと組み合わせてファーストパーティの Cookie 共有を利用して、ユーザーの行動の長期プロファイルを構築することができます。この猫と鼠のゲームでは、トラッカーは個々のユーザーよりも技術的に優位に立っています。

4. ビットからビッグデータまで：追跡ネットワークとはどのようなものか？

あなたをトラッキングするために、ほとんどのトラッキング会社は、ウェブサイトやアプリの開発者を説得して、製品にカスタムトランイングコードを組み込まなければなりません。これは些細なことだとはいえません。トラッキングコードは、パブリッシャー[37]にとって多くの望ましくない効果をもたらす可能性があります。コれはソフトウェアの動作を遅くし、ユーザーを困らせ、GDPR のような法律の下でノ規制を引き起こす可能性があります。しかし、最大のトランキングネットワークは、常に何百万ものさまざまなソースからデータを収集し、膨大な数の Web およびアプリストアをカバーしています。物理的な世界では、トラッカーは看板、小売店、モールの駐車場にひそんでいます。トラッカーはどのように、そしてなぜ広がりをみせているのでしょうか？ このセクションでは、トラッカーのネットワークが実際にどのようなものかについて説明します。

36 訳注： オプトアウト 企業などが個人情報を収集・利用することができるということを事前に決め、本人に知らせておいた上で、後に本人に利用を制限できる機会を与えること。(英辞郎)「企業が個人に行う様々な活動や措置、行為などに対し、対象者がこれを拒否したり、（登録などの）解除・脱退、（情報などの）抹消などを申し出ることをオプトアウトという。
　特に、事前に許諾を得ることなく一方的に行われる電話勧誘やダイレクトメールの配達、電子メール広告の送信などを拒否することや、そのために用意された制度や手続きなどを意味することが多い。国によっては無差別に送信される広告メールに一定の規制を課したり、事業者が勧誘電話を掛けてはいけない電話番号のリストを政府機関などが構築・運営し、消費者からの申し出により登録するといった制度を運用しているところもある。
また、顧客や登録利用者などすでに企業と関わりのある個人が、会員登録の解除や会誌やメールマガジンなどの購読の停止を行うことをオプトアウトということもある。近年では、ネット広告事業者がネット利用者の Web 閲覧履歴をサイトを横断して捕捉するのを拒否したり、企業が取得した個人情報の利用や第三者への提供を拒否することをオプトアウトということもある。」(http://e-words.jp/w/%E3%82%AA%E3%83%97%E3%83%88%E3%82%A2%E3%82%A6%E3%83%88.html)

37 訳注：パブリッシャー ビデオゲーム業界でゲームソフトの販売元（オンラインゲームの場合は運営元）企業のことをパブリッシャーと言ったり、最近ではインターネット業界で Web メディアを所有・運営する事業者のことをパブリッシャー呼ぶ。(http://e-words.jp/w/%E3%83%91%E3%83%96%E3%83%AA%E3%83%83%E3%82%B7%E3%83%A3%E3%83%BC.html)

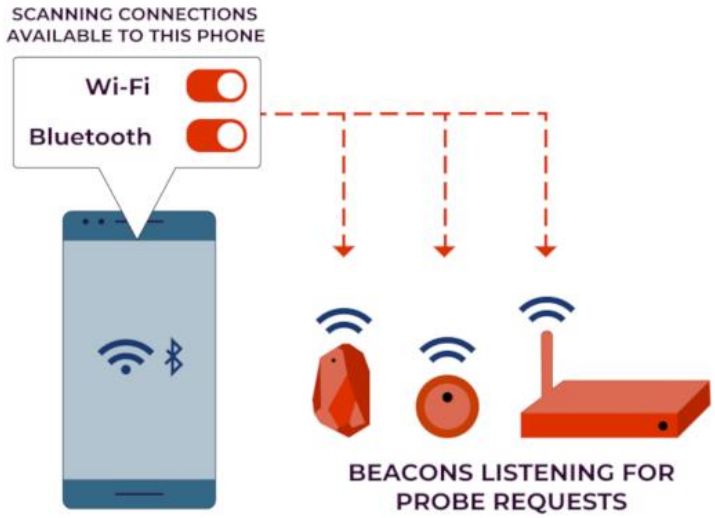

図 4.1: Web 上のトップトラッカー。データを収集する Web トラフィックの割合でランク付けしています。Google は、測定された Web トラフィックの 80% を超えるデータを収集しています。出典：Cliqz GBMH による WhoTracks.me。

4.1. ソフトウェアでの追跡：Web サイトとアプリ

4.1.1. 広告ネットワーク

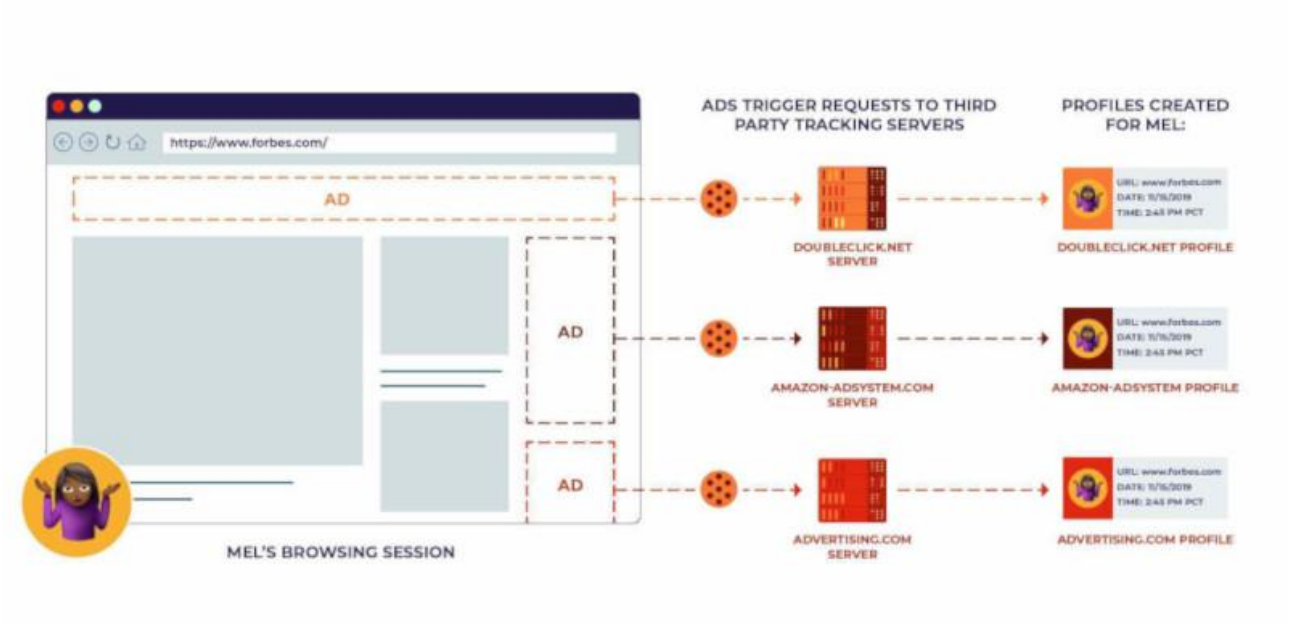

図 4.1.1: あなたのブラウザが読み込む各広告は、異なる広告サーバーからのものである場合があり、各サーバーはアクティビティに基づいて独自のプロファイルを作成することができます。そのサーバーに接続するたびに、Cookie を使用してそのアクティビティをプロファイルにリンクすることができます。

サードパーティのトラッキングの背後にある支配的な市場勢力は、次章で述べるように、

広告業界です。したがって、オンライン広告がデータ収集の主要なベクトルのひとつであることは驚くことではありません。最も単純なモデルでは、単一のサードパーティの広告ネットワークが多数の Web サイトに広告を配信します。広告ネットワークを使用する各サイト運営者は、広告サーバーから広告を読み込むウェブサイトに小さなコードスニペット[38]を含める必要があります。これにより、ユーザーが協力サイトのひとつにアクセスするたびに広告サーバーへのリクエストが引き起こされます。これにより、広告サーバーはユーザーのブラウザにサードパーティの Cookie を設定し、ネットワーク全体のユーザの行動を追跡できます。同様に、広告サーバーは、モバイルアプリ開発者が使用する広告ホスティングソフトウェア開発キット（SDK）を提供する場合があります。ユーザーが SDK を使用するアプリを開くたびに、アプリは広告サーバーにリクエストを送信します。このリクエストにはユーザーのデバイスの広告 ID を含めることができるため、広告サーバーはアプリ全体でユーザーの行動をプロファイリングできます。

実際には、オンライン広告エコシステムはさらに複雑です。アドエクスチェンジは、Web ページ上の個々の広告インプレッションに対して「リアルタイムオークション」を開催します。その過程で、他のいくつかのサードパーティの広告プロバイダーからコードをロードし、オークションに参加している多くの潜在的な広告主と各インプレッションに関するデータを共有する場合があります。表示される各広告は、多数のトラッカーとデータを共有する責任があります。パート 3 では、リアルタイムビッダーやその他のデータ共有アクティビティについてさらに詳しく説明します。

4.1.2. 分析およびトラッキング・ピクセル

多くの場合、トラッキングコードは、サードパーティの広告など、ユーザーに表示されるものとは関連付けられていません。ウェブでは、トラッキングの重要な部分は、トラッカーへのリクエストを引き起すためだけに存在する目に見えない 1 ピクセル x1 ピクセルの「画像」を介して行われます。これらの「トラッキングピクセル」は、Google アナリティクス、Facebook、Amazon、DoubleVerify など、Web 上で最も多くのデータコレクターによって使用されています。[39]

ウェブサイトの所有者がサードパーティのトラッキングピクセルをインストールする場合、通常、サードパーティが収集するデータの一部へのアクセスと引き換えにインストールします。たとえば、Google アナリティクスと Chartbeat はピクセルを使用して情報を収集し、Web サイトの所有者とサイト運営者に、どのような人々がサイトにアクセスしているかについての見解を提供します。さらに深いレベルに進むと、Facebook などの広告プラットフォームは「コンバージョンピクセル」も提供します。これにより、パブリッシャーは、自社のサードパーティ広告が獲得したクリックスルーの数をトラッキングできます。

Web ベースの分析の最大手は、モバイルアプリと同様のサービスを提供しています。

38 訳注： スニペット プログラミングの分野では、様々なプログラムに挿入して利用することができる、特定の機能を実現した短いコードのまとまりのことをコードスニペットあるいは単にスニペットという。(http://e-words.jp/w/%E3%82%B9%E3%83%8B%E3%83%9A%E3%83%83%E3%83%88.html)

39 訳注：Google の説明「インプレッションをトラッキングするには、トラッキング ピクセル コードを使用できます。通常、このコードはサードパーティから URL の形式で入手します。サーバーを呼び出して透明な 1x1 ピクセルを返すコードです。Google ニュースでは、パブリケーション全体または記事ごとにトラッキング ピクセルを適用できます。ユーザーが記事や投稿の最初のページを読み始めると、トラッキング ピクセルの URL が呼び出されます。」(「トラッキング ピクセルを使用する」) Facebook の説明「Facebook ピクセルは、ウェブサイトで実行されたアクションを把握して、広告の効果を測定できる分析ツールです。
ピクセルは次の目的で使用できます。
- 広告が狙いどおりの利用者に表示されているかを確認する。新しい潜在顧客を見つけたり、ウェブサイトで特定のページにアクセスした人や望ましいアクションをとった利用者にリーチしたりできます。
- 販売を促進する。自動入札機能を利用すると、求めているアクション(商品の購入など)を実行する可能性が高い利用者にリーチできます。
- 広告の効果を測定する。利用者が広告を見たときのアクションを測定することで、広告の効果についての理解を深めることができます。」(Facebook ピクセルについて)

カスペルスキー(セキュリティ会社)の記事「ネット上であなたをトラッキングしているのは誰？」も参考になる。

Google Analytics と Facebook は、Android と iOS の両方で最も人気のある SDK の 2 つです。Web 上の同等のサービス同様、これらのサービスはモバイルアプリのユーザーに関する情報を密かに収集し、その情報の一部をアプリ開発者と共有します。

モバイルのサードパーティのトラッカーは、分析やシングルサインオンなどの便利な機能を提供することで、アプリ開発者に SDK をインストールするよう説得します。SDK は、アプリ開発者がプロジェクトに追加するコードの単なる大きなブロブ[40]です。彼らがアプリをコンパイルして配布すると、サードパーティのコードが同梱されます。Web ベースのツールとは異なり、モバイルアプリの分析サービスは、サードパーティのリクエストを引き起こすために「ピクセル」などのトリックを使用する必要がありません。

別のクラスのトラッカーは、ファーストパーティのサイトやアプリとしてよりも、広告主のために機能します。これらの企業は広告主と協力して、広告がどこで、どのように、誰に配信されているかを監視します。多くの場合、ファーストパーティのサイト運営者とはまったく連携しません。彼らの目標は、パブリッシャーとユーザーに関するデータを収集することです。

DoubleVerify は、こうしたサービスのなかでも最大のものひとつです。サードパーティの広告主は、広告と一緒に DoubleVerify コードを挿入し、DoubleVerify は、各インプレッションが（ボットではなく）本物の人間によるものかどうか、人間が広告主のターゲットとする人であるかどうか、広告の周りのページが「ブランドセーフ」であるかどうかを評価します。同社のプライバシーポリシーによると、同社は「広告が消費者のブラウザに表示された時間」と「消費者のブラウザでの広告の表示特性」を測定します。こうすることで、DoubleVerify はユーザーのブラウザに関する詳細なデータを収集すます。これは、Web 上のサードパーティ製ブラウザーのフィンガープリントよりもはるかに大きな最大の資源になります。他のサードパーティのソースからのデータを含む位置データを収集して、広告主がターゲットとする地理的領域の広告をユーザーが表示しているかどうかを判断しようとします。

この分野の他の企業には、 Adobe、Oracle、Comscore が含まれます。

4.1.3. 埋め込みメディアプレーヤー

サードパーティのトラッカーは、ユーザーが実際に見たいコンテンツを提供する場合があります。サードパーティのコンテンツを埋め込むことは、ウェブのブログその他メディアサイトで極めて一般的です。いくつかの例として、
YouTube、Vimeo、Streamable、Twitter などのサービスのビデオプレーヤーや、
Soundcloud、Spotify、Podcast ストリーミングサービスのオーディオウィジェット[41]が含まれます。これらのメディアプレーヤーはほとんど常に IFrame 内で実行されるため、ローカルストレージにアクセスして、任意の JavaScript を実行できます。これにより、ユーザーのトラッキングにも適しています。

4.1.4. ソーシャルメディア・ウィジェット

ソーシャルメディア企業は、Facebook Like ボタンや Twitter Share ボタンなど、さまざまなサービスを Web サイトに提供しています。これらは、多くの場合、パブリッシャーが自分のプラットフォームのトラフィック数とソーシャルメディアでの存在感を向上させる方法として提案されています。 「いいね」ボタンと「共有」ボタンは、ピクセルと同じようにトラッキングに利用できます。「ボタン」は、ソーシャルメディア会社のサーバーへのリクエストを引き起こす埋め込み画像です。

40 訳注： ブロブ(Blob, Binary Large OBject) データベースのフィールド定義などで用いられるデータ型の一つで、テキスト（文字）や数値以外の任意の形式のバイナリデータを格納することができるもの。画像や音声、動画、実行ファイル、圧縮ファイルなど様々な種類のデータをデータベースのレコード中に直接格納するために用いられる。(http://e-words.jp/w/BLOB.html)

41 訳注： ガジェット コンピュータの操作画面を構成する、何らかの機能を持った表示・操作要素。また、それらを組み合わせて作られる、単機能の小さなアプリケーションソフト。後者は「ガジェット」（gadget）とも呼ばれる。(http://e-words.jp/w/%E3%82%A6%E3%82%A3%E3%82%B8%E3%82%A7%E3%83%83%E3%83%88.html)

コメントセクションなどのより洗練されたウィジェットは、埋め込みメディアプレーヤーのように機能します。通常、それらは IFrame の内部にあり、単純なピクセルや画像よりもユーザーのブラウザへのアクセスを楽しんでいます。メディアプレーヤーと同様に、これらのウィジェットはローカルストレージにアクセスし、JavaScript を実行してブラウザーの指紋を計算できます。

最後に、大企業 (特に Facebook および Google) は、「Google でログイン」などの小規模企業にアカウント管理サービスを提供します。「シングルサインオン[42]」として知られるこれらのサービスは、いくつかの理由でパブリッシャーにとっては魅力的です。独立した Web サイトおよびアプリは、ユーザーアカウントの管理作業を大企業に任せることができます。ユーザーは覚えておく必要のあるユーザー名/パスワードのペアが少なくなり、煩わしいサインアップ/ログインフローを実行する頻度が少なくなります。ただし、ユーザーは代償を支払わねばなりません。アカウント管理サービスでは、ログインプロバイダーがサードパーティとして機能し、ログインするすべてのサービスでのユーザーの行動をトラッキングできます。ログインサービスは、ユーザーに ID の確認を強制するため、ピクセルや他の単純なウィジェットよりもより信頼性の高いトラッカーです。

4.1.5. CAPTCHA

CAPTCHA は、ロボットからユーザーを分離しようとする技術です。サイト運営者は、サインアップフォームや特に大きなファイルを提供するページなど、自動化されたトラフィックのブロックに特に注意したいページに CAPTCHA をインストールします。

Google の ReCAPTCHA は、ウェブ上で最も人気のある CAPTCHA テクノロジーです。recaptcha を使用するサイトに接続するたびに、ブラウザは* .google.com ドメインに接続して CAPTCHA リソースをロードし、関連するすべての Cookie を Google と共有します。つまり、CAPTCHA ネットワークは、Google がユーザーをプロファイルするために使用できるデータのもうひとつのソースです。

古い CAPTCHA はユーザーに文字化けしたテキストを読むか、自転車の写真をクリックするように要求しましたが、新しい ReCAPTCHA v3 は「ウェブサイトとの相互作用」を記録し、ユーザーが人間かどうかを密かに推測します。ReCAPTCHA スクリプトは、生のインタラクションデータを Google に送り返しません。むしろ、行動フィンガープリント(behavioral fingerprint)に似たものを生成します。これは、ユーザーがこのページとどのように対話したのかを要約します。Google はこれを機械学習モデルにフィードして、ユーザーが人間である可能性を推定し、そのスコアをファーストパーティの Web サイトに返します。ユーザーにとって便利なことに加えて、この新しいシステムはふたつの点で Google にメリットをもたらします。まず、CAPTCHAS がほとんどのユーザーに表示されないため、Google (または他のユーザー) がユーザーに関するデータを収集していることに気付かない可能性があります。次に、Google の膨大な行動データを活用して CAPTCHA 市場での優位性を強化し、将来競合他社が、同じように機能するツールを構築する独自のインタラクションデータを必要とする場合の防衛策となるものです。

4.1.6. セッション再生サービス

セッション再生サービス(Session replay services)は、ユーザーがサービスとどのようにやり取りするかを記録するために、Web サイトまたはアプリの所有者がインストールできるツールです。これらのサービスは、Web サイトとアプリの両方で作動します。キーストローク、マウスの動き、タップ、スワイプ、およびページへの変更を記録し、その後、

42 訳注： シングルサインオン 一度の利用者認証で複数のコンピュータやソフトウェア、サービスなどを利用できるようにすること。シングルサインオンでは、複数のシステムから横断して利用できる認証基盤を用意し、利用者は一度の認証作業で連携するすべてのシステムにアクセスできるようにする。統合された「唯一の」ID には非常に強力なアクセス権が与えられることになるため、2 段階認証やワンタイムパスワードなど、単純なパスワード認証よりも厳密な認証を求めるようにすることが多い。具体的な ID 連携の技術規格として SAML や OpenID (OpenID Connect) などがある。ID 連携を用いて「Google アカウントでログイン」「Facebook アカウントでログイン」といったように大手ネットサービスのアカウントを流用できるようにしているネットサービスが多数存在する。(http://e-words.jp/w/%E3%82%B7%E3%83%B3%E3%82%B0%E3%83%AB%E3%82%B5%E3%82%A4%E3%83%B3%E3%82%AA%E3%83%B3.html)

ファーストパーティサイトが個々のユーザーエクスペリエンスを「リプレイ」できるようにします。多くの場合、ユーザーは、自分のアクションが記録され、サードパーティと共有されていることを示されることはありません。

こうした不気味なツールは、医療情報、クレジットカード番号、パスワードなどの機密データが記録され、漏洩される大きなリスクを生み出します。通常、セッションリプレイサービスのプロバイダーは、特定のデータを立ち入り禁止に指定することをクライアントに任せます。しかし、クライアントにとって、機密情報を除外するプロセスはややこしく、骨が折れ、時間がかかり、「数秒で」セットアップするというリプレイサービスの約束とは衝突します。その結果、独立監査は、機密データが記録されされていること、セッション再生サービスプロバイダーはそのデータをしばしば適切に保護できていないことを発見してきました。

4.2. 受動的な実世界の追跡

4.2.1. WiFi ホットスポットとワイヤレスビーコン

多くの民生機器は無線「プローブ」信号を発信し、多くの企業はこれらのプローブを物理世界全体で傍受する商用ビーコンをインストールします。一部のデバイスは、プローブで共有する一意の MAC アドレスデバイス識別子をランダム化し、パッシブトラッキングから自身を保護しますが、すべてがそうとは限りません。また、オープンな WiFi ネットワークに接続するか、アプリに Bluetooth の許可を与えると、常にこうしたデバイスを追跡できるようになります。

上記で説明したように、WiFi ホットスポット、ワイヤレスビーコン、およびその他の無線デバイスを使用して、近くのデバイスを「盗み聞き」できます。 Comcast (XFinity WiFi を提供) や Google (スターバックスやその他の多くの企業で無料の WiFi を提供) などの企業は、世界中に WiFi ホットスポットを設置しています。Comcast だけでも、1800 万を超える XFinity WiFi がインストールされています。聞いたことがないと思われる他の多くの企業は、コーヒーショップ、レストラン、イベント、ホテルに無料の WiFi を提供しています。

企業はまた、現実のビジネスや公共スペースにワイヤレスビーコンを設置するための投資もします。Bluetooth 対応ビーコンは、小売店の周り、政治集会、キャンペーンの芝生看板、街灯に設置されています。

ワイヤレスビーコンは、ふたつのレベルでトラッキングできます。まず、最も懸念されるのは、ワイヤレスビーコンが、デバイスが常に送信する「プローブ」を受動的に監視できることです。デバイスがハードウェア MAC アドレスをブロードキャストしている場合、企業は収集したプローブを使用して、ユーザーの経時的な動きを追跡できます。

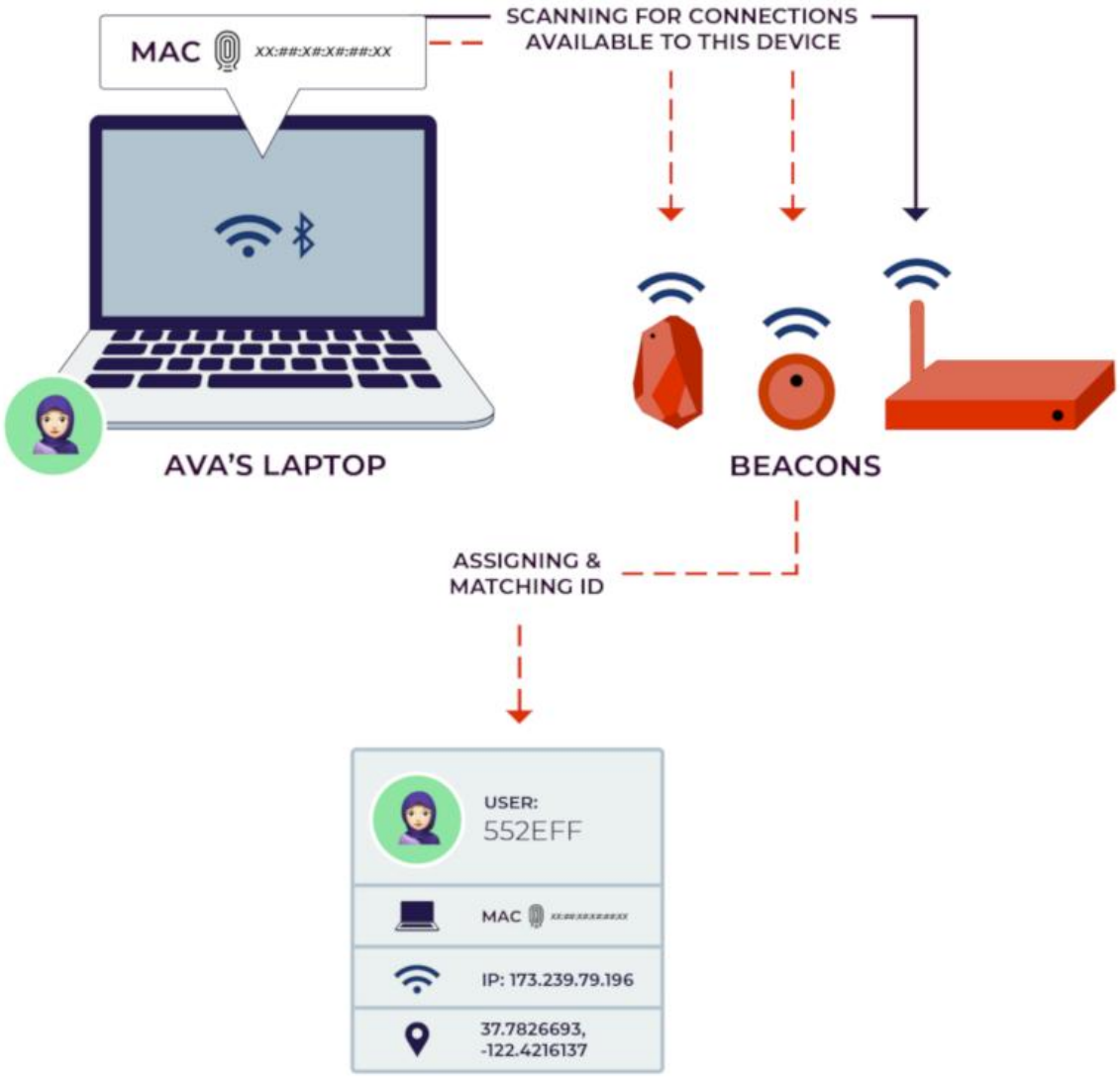

図 4.2.1: WiFi ホットスポットと Bluetooth ビーコンは、ワイヤレスデバイスが自動的に送信するプローブを聞くことができます。 トラッカーは、各デバイスの MAC アドレスを使用して、そのデバイスを見た場所に基づいてプロファイルを作成できます。

次に、ユーザーが WiFi ホットスポットまたは Bluetooth ビーコンに接続すると、ホットスポットまたはビーコンのコントローラーは、デバイスの MAC アドレスを IP アドレス、クッキー、広告 ID などの追加の識別子に接続できます。多くの WiFi ホットスポットオペレーターは、サインインページを使用して、ユーザーの本名またはメールアドレスに関する情報を収集します。その後、ユーザーがそのホットスポットからウェブを閲覧すると、オペレーターは ISP のように、ユーザーのデバイスからのすべてのトラフィックに関するデータを収集できます。Bluetooth ビーコンの利用方法は少し異なります。携帯電話を使用すると、アプリは特定の権限で Bluetooth インターフェイスにアクセスできます。Bluetooth 権限を持つアプリのサードパーティのトラッカーは、現実世界の Bluetooth ビーコンに自動的に接続でき、それらの接続を使用して、きめの細かい位置データを収集できます。

ありがたいことに、iOS デバイスと Android デバイスの両方が、デフォルトで難読化された MAC アドレスをプローブで送信します。これにより、上記の最初の受動的な追跡スタイ

ルは妨げられます。

しかし、ワイヤレス機能を備えたデバイスは電話だけではありません。ラップトップ、電子ブックリーダー、ワイヤレスヘッドフォン、さらには車にさえ、Bluetooth 機能が装備されていることがよくあります。これらのデバイスの一部には、最近のスマートフォンのモデルが持つ MAC ランダム化機能がないため、受動的な位置トラッキングに対して脆弱です。

さらに、MAC ランダム化を備えたデバイスでさえ、通常、実際にワイヤレスホットスポットまたは Bluetooth デバイスに接続するときに静的 MAC アドレスを共有します。これにより、デバイスが公衆 WiFi ネットワークまたはローカル Bluetooth ビーコンに接続するときに発生する上記の 2 番目の追跡のリスクが高まります。

4.2.2. 車両追跡と ALPR

自動化されたナンバープレート読み取り機（ALPR）は、ナンバープレートを検出して読み取る機能を備えたカメラです。また、この識別を助けるために、メーカー、モデル、色、および摩耗のようなその他の車の特性を利用することができます。ALPR はしばしば法執行機関によって使用されますが、多くの ALPR デバイスは民間企業が所有しています。これらの企業は無差別に車両データを収集し、入手した後は、地元の警察、連邦移民法執行機関、個人データ収集業者、保険会社、貸し手、または賞金稼ぎなど、希望する人に再販売できます。

さまざまな企業がさまざまなソースからナンバープレートデータを収集し、さまざまな対象者に販売しています。デジタル認識ネットワーク（DRN）は、全国の何千もの差し押さえ機関からデータを入手し、保険機関、民間調査員、および「資産回収」会社にデータを販売しています。Motherboard による調査によると、DRN がデータを収集する個人の大多数は、犯罪の疑いや車の支払いの滞納者といった人たちではありません。新興企業の Flock Safety は、ALPR で駆動する監視サービスを提供しています。関心をもつ住宅所有者は、自分の所有物に ALPR をインストールして、近所を走る車の情報を記録・共有できてしまいます。

DRN は、フォートワースに本拠を置く VaaS International Holdings が所有し、「ナンバープレート認識（ 'LPR'）および顔認識製品およびデータソリューションの卓越したプロバイダー」というブランドを誇っています。もうひとつの ALPR テクノロジーの民間プロバイダー、Vigilant Splitions のクライアントには、法執行機関や商業ショッピングセンターが含まれています。Vigilant は、全国の数千のソースからのデータを「PlateSearch」と呼ばれる単一のデータベースにプールします。多数の法執行機関が PlateSearch へのアクセスに対して支払いをしています。EFF の調査によると、Vigilant によって記録されたナンバープレートの約 99.5%は、スキャンされた時点で公共の安全とは無関係なものです。

車両追跡を可能にする技術は、カメラとマシンビジョンだけではありません。パッシブ MAC アドレスの追跡は、車両の動きの追跡にも使用できます。車両内の電話、および場合によっては車両自体が、MAC アドレスを含むプローブ要求をブロードキャストします。道路の周囲に戦略的に配置されたワイヤレスビーコンは、それらの信号を傍受することができます。Libelium という会社は、近くのトラフィックを追跡するために街灯に設置する目的でワイヤレスビーコンを販売しています。

4.2.3. 顔認識カメラ

顔認証は、中国や英国を含む一部の国で法執行機関によって広範囲に展開されています。これは恐ろしいことです。無実の人々の活動を大量に記録できるのです。中国では、ウイグル族のマイノリティのコミュニティの人々を監視・管理するのに利用されています。

法執行機関による顔認証の使用による市民的自由の侵害に関しては、これまでも幅広く取り上げられてきました。しかし、顔認識は多くの民間産業でも展開されています。航空会社は、搭乗前に顔認証を使用して乗客を認証します。コンサート会場とチケット売り手

は、コンサートの観客を選別するために利用してきました。小売業者は、顔認識を利用して、万引きリスクが高いと疑われる人々を特定しています。これは、特に、基礎となる逮捕者顔写真データベース(マグショットデータベース)が不公平な人種格差にむしばまれており、技術的に有色人種を誤認する可能性が高いことを考えると危惧すべきことです。民間警備会社は、顔認識機能を備えたロボットを販売して公共スペースを監視したり、雇用主が従業員を監視できるようにしています。学校やサマーキャンプでさえ、子どもたちを監視するために利用しています。

大手テクノロジー企業は、支払い処理のために顔認識に投資し始めました。これにより、現実世界の行動をユーザーのオンライン上の人格にリンクするもうひとつの方法が提供されます。Facebook は、支払い処理のために、顔をソーシャルメディアプロファイルにリンクするシステムで特許を申請しました。また、Amazon の実店舗の「Go」ストアは、生体認証に依存して、誰が入場し、何を買ったのかを追跡し、それに応じた課金をします。

さらに、多くの人たちは、顔認識を物理的な世界にターゲット広告をもちこむ論理的な方法だと考えています。顔認識カメラは、店舗、看板、ショッピングモールに設置されて、人々の行動をプロファイリングし、習慣に関する書類を作成し、人々に対するターゲットメッセージを作成することができます。2019 年 1 月、Walgreens は冷蔵庫のドアに LED スクリーンを装備した顔認識カメラを使用したパイロットプログラムを開始しました。このアイディアでは、消費者は、冷蔵庫の中身を見るためにガラスの板を覗くのではなく、中に何が入っているかを示すグラフィック表示画面を見るのです。カメラは冷蔵庫の前にいる人の顔を認識でき、グラフィックはその人をターゲットにした広告を配信するように動的に変更することができます。Walgreens がこの技術を大規模に展開するかどうかにかかわらず、これは小売業者が向かう方向の 1 つであると思われます。

4.2.4. 支払い処理業者と金融技術

金融テクノロジー、または「フィンテック」は、急成長する金融隣接テクノロジー企業の業界を包括する用語です。数千の比較的新しいハイテク企業が、旧来の金融機関とトラッキングと監視を含む新しい技術との間の技術的接着剤として機能しています。規制される場合でも、フィンテック企業は多くの場合、銀行などの従来の機関よりも政府による監督は小さいのです。

支払い処理業者は、他のビジネスに代わって支払いを受け入れる会社のことです。彼らは企業が何を売り、人々が何を買うかについての膨大な情報を持っています。ほとんどの金融取引にはクレジットカード番号と名前が含まれるため、支払い処理業者は収集したデータを実際の身元に簡単に結び付けることができます。これらの会社の一部は純粋なサービスプロバイダーであり、ある場所から別の場所にお金を移動する以外の目的でデータを使用しません。しかし他の業者は、消費者または企業のプロファイルを作成し、そのデータを収益化します。たとえば、Square は中小企業向けのクレジットカードリーダーを製造する会社で、第三者からターゲット広告を収集し、この企業の Square Capital プログラムを通じてローンを引き受けるために収集した情報を利用します。

TurboTax と Mint の背後にある Intuit という会社のように、一部のフィンテック企業はユーザーに金融サービスを直接提供しています。その他は、銀行または企業にサービスを提供します。フィンテックの世界では、「データアグリゲーター」は銀行と他のサービス（資金管理アプリなど）の仲介役として機能します。この過程で、データアグリゲーターは、数百万人の口座残高、未払い債務、クレジットカードトランザクションなど、パイプを通過するすべてのデータにアクセスできます。さらに、アグリゲーターは多くの場合、消費者のユーザー名とパスワードを収集して、銀行からデータを抽出します。この分野で最大の企業のひとつである Yodlee は、ヘッジファンドに取引データを販売し、ヘッジファンドは情報をマイニングして株式市場の動きを知らせます。多くのユーザーは、サインアップしたアプリの操作以外にデータが使用されていることに気づいていません。

4.3. 追跡と企業力

データ追跡の恩恵を最も受けている企業の多くは、Web 開発者、アプリ作成者、ストアマ

ネージャーにトラッキング技術をインストールするよう促す魅力的な方法を持っています。独占企業または準独占企業は、市場力を駆使して追跡ネットワークを構築し、小規模な競合他社を監視あるいは阻止し、消費者のプライバシーを悪用して経済的利益を得ることができます。企業力と企業の監視は、いくつかの方法で相互に補強しあいます。

まず、Google や Facebook などの支配的な企業は、パブリッシャーにトラッキングコードのインストールを強制することができます。サイト運営者は、世界最大のソーシャルネットワークと世界最大の検索エンジンを使用して、自分のサイトにトラフィックを誘導します。その結果、ほとんどのパブリッシャーはこれらのプラットフォームで広告する必要を生じます。また、広告の効果を追跡するには、サイトとアプリに Google と Facebook のコンバージョン測定コードをインストールするほかありません。Google、Facebook、および Amazon もサードパーティの広告ネットワークとして機能し、これら合わせて市場の 3 分の 2 以上を支配しています。つまり、コンテンツを収益化しようとするサイト運営者は、大手プラットフォームの広告トラッキングコードを回避するのに苦慮するのです。

第二に、垂直統合されたハイテク企業は、トラッキング市場の両側を支配できます。Google は世界最大の行動ターゲティング広告システムを管理しており、世界で最も人気のあるモバイルオペレーティングシステムの Android スマートフォンと最も人気のある Web ブラウザーである Chrome ブラウザーからデータを収集しつつ機能しています。Google のユーザーソフトウェアは、他のオペレーティングシステムやブラウザと比較して、トラッカーにとってデータ収集がしやすくなっているのです。

Web の設計者が初めてブラウザについて説明したとき、彼らはこれを「ユーザーエージェント」 と呼びました。インターネット上でユーザーの代わりに動作するソフトウェアという意味です。しかし、ブラウザメーカーが行動ターゲティング広告を主な収入源とする企業でもある場合、ユーザーのプライバシーと制御に対する企業の利害と、トラッキングに対する企業の利害とが相反することになります。通常、会社の収益が最優先されます。

第三に、データは人々だけでなく競合他社のプロファイルにも利用できます。最大のデータコレクターは、私たちがどのように行動するかを知っているだけでなく、他の誰よりも市場と競合他社についても知っています。Google の追跡ツールは、ウェブ上のトラフィックの 80%以上を監視しています。つまり、競合他社のトラフィック（またはそれ以上）と同じくらい多くのトラフィックを把握していることがよくあります。Facebook（サードパーティの広告、アナリティクス、コンバージョンピクセル、ソーシャルウィジェット、以前の VPN アプリ Onavo を介して）も、大小の Web サイト、アプリ、パブリッシャーの使用と成長を監視します。 Amazon はすでに Amazon Web Services コンピューティングクラウドでインターネットの大部分をホストしており、独自の強力なサードパーティの広告ネットワークを構築し始めています。これらの巨人はこの情報を利用して新手の競合他社を特定し、重大な脅威になる前にこれらを買収するか、製品のクローンを作成します。内部の機密文書によると、Facebook は WhatsApp の買収を告知するために、Onavo やその VPN からユーザーのアプリの習性に関するデータを利用しました。[43]

第四に、技術の巨人は追跡力を自分の手に集中させるために、データへのアクセスを反競争的な武器として使用できます。Facebook は、その API[44]（およびそれに伴う詳細なプライベートデータ）へのアクセスが他のソーシャル企業にとって非常に貴重であることを十分に認識していました。競争力を削ぐために、ユーザーデータへのアクセスを許可または

43 訳注：Onavo 「2013 年に Facebook が推定 1 億ドル(約 110 億円)以上で買収した Onavo は、プライバシー保護をうたうモバイル端末向け VPN アプリ「Onavo Protect」をリリースしています。ところが、「Onavo Protect はユーザー情報を収集して Facebook に送信している」という事実が明らかになり、Onavo Protect が App Store から削除される事態になりました。」（「Facebook の VPN アプリ「Onavo Protect」が App Store から削除される」https://gigazine.net/news/20180823-onavo-vpn-app-remove/）

44 訳注： API アプリケーションプログラミングインターフェース。あるコンピュータプログラム（ソフトウェア）の機能や管理するデータなどを、外部の他のプログラムから呼び出して利用するための手順やデータ形式などを定めた規約のこと。個々のソフトウェアの開発者が毎回すべての機能をゼロから開発するのは困難で無駄なため、多くのソフトウェアが共通して利用する機能は、OS やミドルウェアなどの形でまとめて提供されている。
そのような汎用的な機能を呼び出して利用するための手続きを定めたものが API で、個々の開発者は API に従って機能を呼び出す短いコードを記述するだけで、自分で一から処理内容を記述しなくてもその機能を利用したソフトウェアを作成することができる。(http://e-words.jp/w/API.html)

差し控えるという記憶すべき過去があります。

さらに、Google と Facebook はどちらも、自分たちは無制限にデータを収集しながら、彼らのデータに競合他社がアクセスするのを制限するするポリシーを採用し始めました。たとえば、大規模なプラットフォームのほとんどは、独自のサイトでサードパーティのトラッカーを制限しています。独自の RTB[45]バージョンでは、Google は最近、競合する広告ネットワークがユーザープロフィールを作成できるようにする広告識別子その他の情報へのアクセスを制限し始めました。そして、ケンブリッジアナリティカの事件の後、Facebook はサードパーティ API へのアクセスのロックダウンを開始しました。Facebook 自体がユーザーに関して収集するデータに関して有意義な変更は一切ありません。一方で、サードパーティのアクセスを制限すると、プライバシー上の利点が得られます。他方で、サードパーティのデベロッパーや外部のアクターを Facebook や Google のプラットフォームサービスから追い出すことで、競争上の問題を悪化させ、既存の巨人はかれらが収集したユーザーデータへの独占的な権限が与えられ、プライバシーに関して有害なビジネス慣行を固定化することを可能にしてしまいます。競争とプライバシーとを別々の独立した懸念と見なすのではなく、ユーザーに権限を与えるには、大企業がユーザーのデータへの支配と関心を軽減するために、この両方に対処する必要があります。

最後に、大企業は合併や買収で他の企業から大量のデータを取得できます。Google アナリティクスは、2005 年に Google が買収した独立企業 Urchin から誕生しました。2007 年、Google は Doubleclick を買収してサードパーティの広告ネットワークを強化し、現在は行動ターゲティング広告市場のリーダーとなっています。2019 年後半には、健康データ会社 Fitbit を買収し、長年にわたる歩数や運動ログをユーザーの身体活動に関する膨大な独自のデータベースに統合しました。

Facebook は短期間に、67 社を買収しました。Amazon は 91 社を、Google は 214 社を獲得しています。これは年間平均 10 以上の数になります。Facebook、Amazon、または Google が買収した小規模企業の多くは、膨大なデータと数百万人のアクティブユーザーへのアクセスを有していました。買収のたびに、これらのデータソースは、テクノロジーの巨人が支配する大規模なサイロに投入されることになります。また、ネットワーク効果によって、データがすべてひとつの屋根の下にある場合に、データの価値は高まります。Doubleclick は単独で、ユーザーの閲覧履歴の仮名プロファイルを作成できました。Google の一部となると、このデータを実際の名前、場所、クロスデバイスアクティビティ、検索履歴、ソーシャルグラフとマージすることができるようになります。

私たちを追跡している企業は数十億ドルの巨人だけではありませんし、世界で最も無責任な存在だというわけでもありません。しかし、彼らが巨大であればあるほど、彼らはより多くを知る立場に立つことになります。また、企業がアクセスするデータの種類が増えるほど、ユーザーと競合他社に関するプロファイルはより強力なものになります。個人情報の新しい経済では、金持ちだけがますます金持ちになっていきます。

5．データ共有：ターゲティング、ブローカー、リアルタイム入札

収集されたデータはどこに行くのでしょうか？ほとんどのトラッカーは、すべての情報を自分で収集するわけではありません。代わりに、企業はお互いに協力してデータを収集し、相互に共有します。時には、同じ個人に関する情報を持つ複数の企業は、これを簡単に組み合わせて、その人に対して、どの広告主がどのような広告を配信するかのを決定します。そのほかの場合として、企業は、これまで関わりをもたなかった個人に関するデータの収集と販売に基づいてビジネスモデル全体を構築します。どの場合においても、彼らが収集し共有するデータのタイプは、彼らがさらされている広告に影響を与えるとか、最終的にカタログ化される政府データベースを決定するとかを通じて、対象となっている人の経験に影響を与え、有害なデータへの侵害リスクが大きくなります。このセクションでは、個人情報がどのように共有されるのか、そしてどこに行くのかについて説明します。

45 訳注：RTB リアルタイムビッディング ネット広告の取引手法の一つで、広告が表示される瞬間に複数の広告代理店や広告主で入札を行い、最も条件の良い広告を掲載する方式。(http://e-words.jp/w/RTB.html)

5.1. リアルタイム入札

リアルタイム入札は、サイト運営者と広告主がターゲット広告を配信するために使用するシステムです。インターネット広告の世界での販売単位は「インプレッション」です。広告を掲載しているWebページに人がアクセスするたびに、その人は広告のインプレッションを見るのです。舞台裏では、広告主はあなたに広告の表示権に広告ネットワークに対して支払い、広告ネットワークはあなたが広告を見たウェブページのパブリッシャーに支払います。しかし、こうしたことが起こる前に、パブリッシャーと広告ネットワークはどの広告を表示するノかを決定しなければなりません。このために、彼らは、ミリ秒単位のオークションを実施します。このオークションでは、競売人がユーザーの個人情報を提供し、その後、多数の企業サーバー上のソフトウェアがそのユーザーの注意を引く権利に入札します。データは一方向に流れ、お金は逆方向に流れます。

このような「リアルタイム入札」は非常に複雑であり、このトピックだけで単独の白書が必要なくらいです。幸いなことに、このトピックに関する膨大で詳細な資料が既にあります。ジョニーライアン博士Dr. Johnny Ryanとブレイブ Braveは、RTBのプライバシーへの影響に関する一連の論文を書きました。プロトコルのプライバシーへの影響に関する博士論文もあります。このセクションでは、プロセスの概要を簡単に説明します。その多くは、ライアンの研究に基づいています。

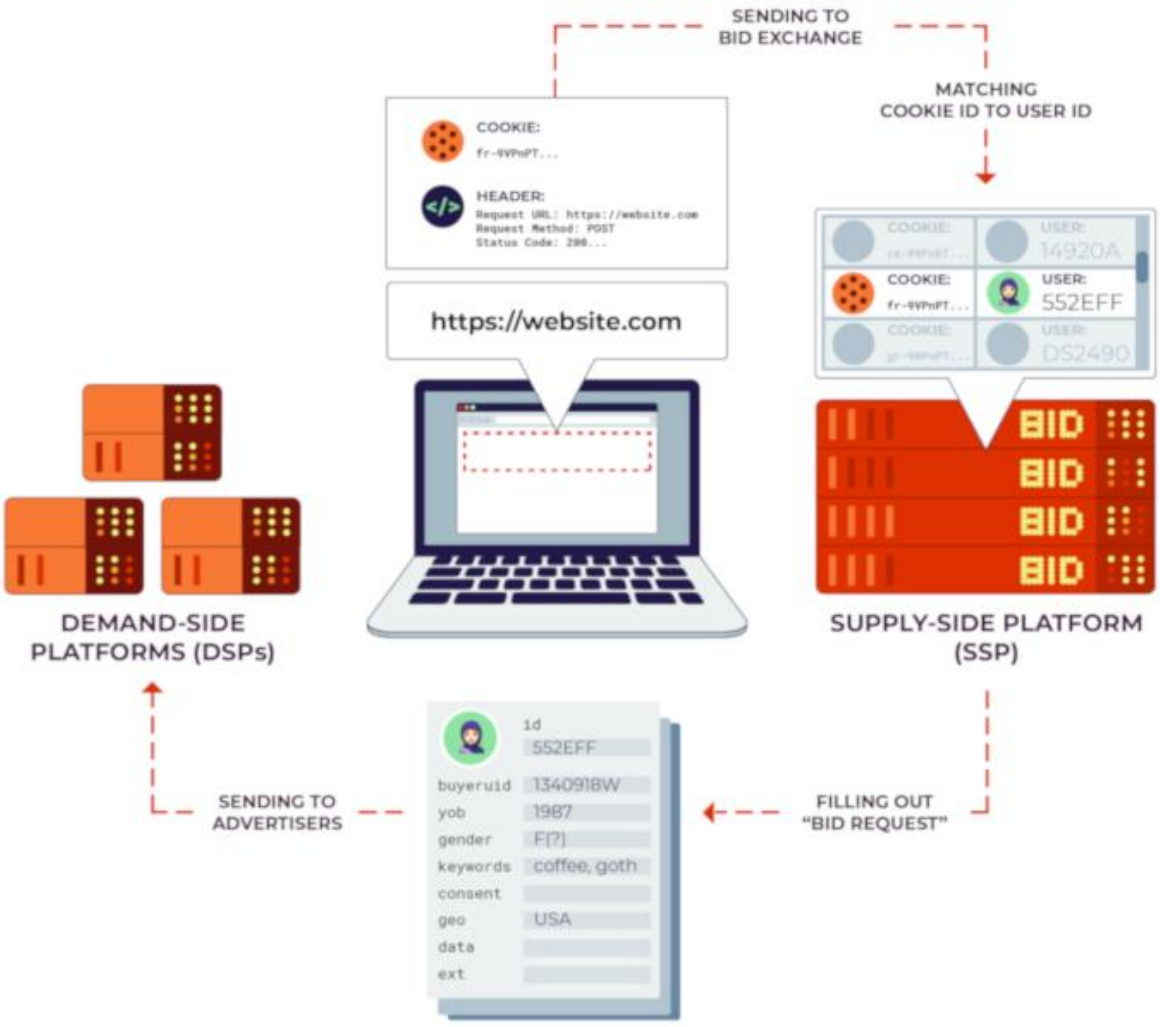

図 5.1.1: サプライサイドプラットフォームは、Cookieを使用してユーザーを識別し、ユーザーに関する情報とともに「入札リクエスト」を潜在的な広告主に配信します。

まず、データはブラウザから「サプライサイドプラットフォーム（SSP）」[46]とも呼ばれる広告ネットワークに流れます。この経済では、あなたのデータとあなたの関心が広告ネットワークと SSP によって販売され、これが「供給」になりす。各 SSP は、通常は Cookie の形式であなたの識別情報を受信し、あなたの過去の行動について知っていることに基づいて「入札リクエスト」を生成します。次に、SSP はこの入札リクエストを、広告の表示に関心を示した数十の広告主のそれぞれに送信します。

Attribute	Type	Definition
`id`	string; recommended	Vendor-specific ID for the user. At least one of `id` or `buyeruid` is strongly recommended.
`buyeruid`	string; recommended	Buyer-specific ID for the user as mapped by an exchange for the buyer. At least one of `id` or `buyeruid` is strongly recommended.
`yob`	integer	Year of birth as a 4-digit integer.
`gender`	string	Gender, where “M” = male, “F” = female, “O” = known to be other (i.e., omitted is unknown).
`keywords`	string	Comma separated list of keywords, interests, or intent.
`consent`	string	GDPR consent string if applicable, complying with the comply with the IAB standard Consent String Format in the Transparency and Consent Framework technical specifications.
`geo`	object	Location of the user's home base (i.e., not necessarily their current location). Refer to Object: Geo.
`data`	object array	Additional user data. Each `Data` object represents a different data source. Refer to Object: Data.
`ext`	object	Optional vendor-specific extensions.

図 5.1.2: OpenRTB 入札リクエストの「ユーザー」オブジェクトには、1 つ以上の一意の ID、年齢、性別、場所、興味など、特定の供給側プラットフォームがインプレッションの主題について知っている情報が含まれています。 ソース：https://github.com/InteractiveAdvertisingBureau/AdCOM/blob/master/AdCOM%20v1.0%20FINAL.md#object--user-

46 訳注：サプライサイドプラットフォーム インターネット広告の配信支援ツールの一。広告枠をもつ配信事業者側の利益が最大となるよう、最適な広告の選択と広告枠の販売を支援する。広告の選択および販売額の決定はアドエクスチェンジとリアルタイム入札を組み合わせて自動的に行われる。(https://kotobank.jp/word/%E3%82%B5%E3%83%97%E3%83%A9%E3%82%A4%E3%82%B5%E3%82%A4%E3%83%89%E3%83%97%E3%83%A9%E3%83%83%E3%83%88%E3%83%9B%E3%83%BC%E3%83%A0-1720784)
　オンライン広告において、媒体社（メディア）の広告枠の販売や広告収益の最大化などを支援するツールのことである。略して「SSP」と呼ばれることが多い。
　サプライサイドプラットフォームでは、主に、広告のインプレッションが発生するごとに最適な広告を自動的に選択し、収益性の向上を図る、といった仕組みが提供されている。具体的にどのような方式が採用されているかは個々のサービスによって異なるが、アドネットワークの一元的管理、アドエクスチェンジ、リアルタイム入札（RTB）、などに対応している場合が多い。
　配信する広告が決定される際には、eCPM だけではなく、ユーザーの行動履歴や属性などといったオーディエンスデータなどが利用される場合が多い。また付加的な機能として、特定の枠に純広を配信し、配信期間やインプレッション回数を制御するといった、アドサーバーとしての利用も可能である場合が多い。これらによって収益最大化だけでなく広告管理の全体的な効率化が実現できる。
　2011 年 11 月現在、国内で提供されているサプライサイドプラットフォームの例としては、オーディエンスターゲティングを中心とした MicroAd の「ADfunnel」、訪問者の広告閲覧回数に基づいた広告収益の最大化を特長とした「Kauli」、スマートフォン向けのサプライサイドプラットフォーム「SSPapri」などがある。
　なお、媒体社の側を支援するサプライサイドプラットフォームに対して、クライアント側（広告主）を支援するシステムは、「デマンドサイドプラットフォーム」と呼ばれている。(https://www.sophia-it.com/content/Supply-Side+Platform)

入札リクエストには、現在地、興味、デバイスに関する情報が含まれ、一意の ID が含まれます。 上記のスクリーンショットは、OpenRTB 入札リクエストに含まれる情報を示しています。

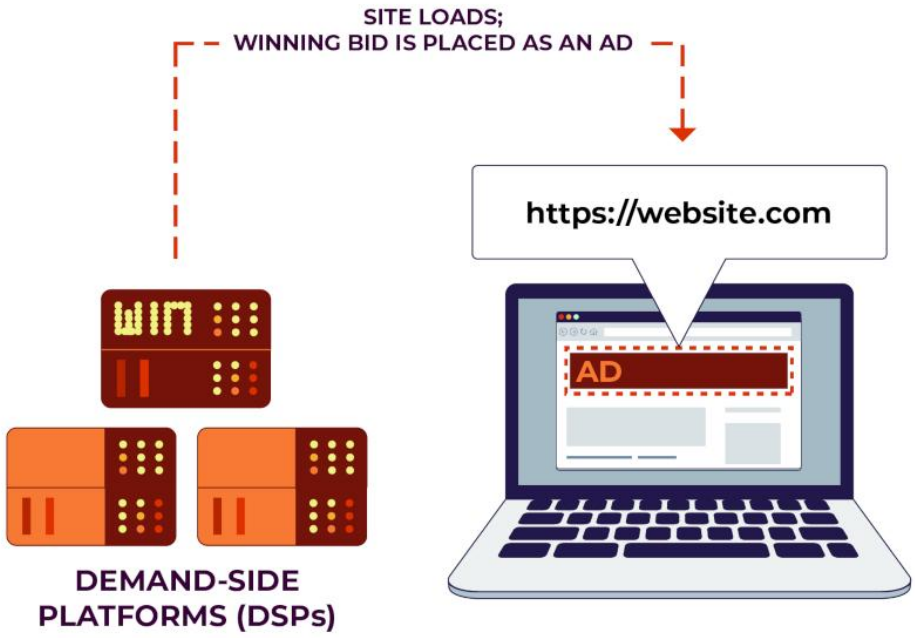

図 5.1.3: オークションの完了後、落札者はサプライサイドプラットフォームに支払い、SSP はパブリッシャーに支払い、パブリッシャーはユーザーに広告を表示します。 この時点で、勝者の広告主はユーザーのブラウザからさらに多くの情報を収集できます。

最後に、入札者の番です。 自動化されたシステムを使用して、広告主はあなたの情報を見て、あなたに広告を出したいかどうか、どの広告を表示したいかを決定し、入札で SSP に応答します。SSP は、誰がオークションに勝ったかを判断し、パブリッシャーの Web ページに勝者の広告を表示します。

入札リクエストのすべての情報は、お金の支払いが行なわれる前に共有されます。オークションに勝てない広告主は、引き続きユーザーの個人情報を受け取ります。これにより、「シャドウ入札」が可能になります。特定の企業は、インプレッションの購入に興味を持っているふりをしますが、各オークションで意図的に落札しないでできるだけ多くのデータをできるだけ安く収集することを目的するものもあります。

さらに、SSP と広告主の間には RTB に参加する企業のいくつかの階層があり、企業の各層もユーザー情報を吸い上げます。「アドエクスチェンジ」[47]をもつ SSP は、データを「デマンドサイドプラットフォーム」(DSP)[48]と共有し、また、データブローカーとデータを共

47 訳注： アドエクスチェンジ アドエクスチェンジとは、広告枠をインプレッションベースで取引する広告取引市場のことです。広告主側の需要とメディア側の供給のバランスにより、インプレッションごとに広告枠の価格が決まります。
アドネットワークでは「クリック課金型」「インプレッション課金型」など、ネットワークにより課金形態が異なることも多いですが、アドエクスチェンジという広告取引市場を介することによって、広告主から見ると“入札方式のインプレッション課金型“で統一されました（広告の取引市場化）。(https://dmlab.jp/words/d002.html)

48 訳注： デマンドサイドプラットフォーム(DSP) Web 広告の配信仲介システムの一種で、広告主の希望する条件に基いて、複数のメディアやアドネットワークなどに一括して広告を出稿できるもの。また、そのようなシステムを広告主や広告代理店に提供するサービス。

有したり買入れたりします。パブリッシャーは彼らの広告スペースを販売するために SSP と協力し、広告主はこれを買うために DSP と協力し、広告の交換は売り手と買い手を結びつけます。広告主向けに書かれた SSP と DSP の違いの内訳はここで読むことができます。このプロセスに関与する全員が、リクエストを引き起こした人物に関する行動データを収集できます。

入札プロセスの間、広告主と提携する DSP は、サードパーティのデータブローカーを利用して、個々のユーザーのプロファイルを強化することができます。これらのデータブローカーは、自らを「データ・マネジメント・プラットフォーム」（DMP）[49]と称し、入札リクエストに含まれる識別子と人口統計に基づく個人に関するデータを販売しています。つまり、広告主はユーザー ID をデータブローカーと共有し、そのユーザーの行動プロファイルを受け取ることができるのです。

出典：Zhang, W., Yuan, S., Wang, J., and Shen, X. (2014b). Real-time bidding benchmarking with ipinyou dataset. arXiv preprint arXiv:1407.7073.

上の図の単一の RTB オークションでの情報とお金の流れをもう一度見てみます。

要約すると：（1）ユーザーがページにアクセスすると、ページのパブリッシャーからアドエクスチェンジに広告リクエストが出されます。これがリアルタイム入札の「競売人」です。アドエクスチェンジ（2）は、広告主とその広告主と連携する DSP に入札を要求し、処理中のユーザーに関する情報を送信します。DSP は、（3）データブローカーまたは DMP からの詳細情報を使用して入札リクエストデータを補強します。広告主（4）は、広告スペースの入札で応答します。（5）ミリ秒のオークションの後、アドエクスチェンジ（6）が落札者を選び、通知します。アドエクスチェンジ（7）は、上記のトラッキングテクノロジーを備えたユーザーにその広告を配信します。広告主は（8）ユーザーが広告をどのように操作したかに関する情報を受け取ります。閲覧時間、クリックしたもの、購入したものなど。そのデータは、そのユーザーと特性を共有する他のユーザーに関する DSP の情報にフィードバックされ、将来の RTB 入札を通知します。

ページを訪問したユーザーの観点から、RTB は 2 つの別個のプライバシー侵害を引き起こします。まず、ページにアクセスする前に、多数の企業がオンラインとオフラインの両方で個人情報を追跡し、すべてをそれらに関する洗練されたプロファイルにマージしています。次に、RTB プロセス中に、異なる複数の企業がこのプロファイルを利用して、広告インプレッションの入札額を決定します。第二に、ユーザーがページにアクセスした結果、RTB 参加者は訪問ユーザーから追加情報を収集します。その情報はユーザーの古いプロ

DSP に広告の掲載を依頼すると様々なメディアやアドネットワーク、SSP が行っている広告枠の RTB（リアルタイム入札）に参加し、落札した掲載枠に広告クリエイティブを配信・掲載する。
配信数や落札単価、クリックなど掲載後の効果はすべて記録され、キャンペーン毎やクリエイティブ毎、掲載メディア毎、掲載枠毎などの単位で集計・分析することができる。分析結果に基いて、最適な配信ができるよう配信先や単価などを自動的に調節する機能もある。(http://e-words.jp/w/DSP-2.html)

49 訳注： データ・マネジメント・プラットフォーム(DMP) インターネット上の様々なサーバーに蓄積されるビッグデータや自社サイトのログデータなどを一元管理、分析し、最終的に広告配信などのアクションプランの最適化を実現するためのプラットフォームのことです。
DMP は大きく 2 種類に分類されます。1 つは広告配信先のデータセラーとしての機能を果たすタイプの「オープン DMP」、もう 1 つは企業が自社で蓄積した Web ログや顧客 DB などの蓄積したデータを利用するタイプの「プライベート DMP」です。
「オープン DMP」は、Web サイト訪問ユーザーのデモグラ情報や、興味関心・嗜好性等などの外部のオーディエンスデータとデータエクスチェンジさせることができるクラウド型のデータプラットフォーム（様々な Web サイトのオーディエンスデータを集約して整理するデータ格納庫のようなもの）のことです。
「プライベート DMP」は、オープン DMP の領域に加え、企業独自のマーケティングデータ（購買情報、ユーザープロファイル、各種プロモーションの結果等）を集約し、それを外部のオーディエンス情報とシンクさせ構築するプラットフォームです。CRM データに従来では取得することが難しかった外部データを組み合わせたものだと理解しやすいと思います。データ格納先が企業側にあるというところがポイントです。
こういった説明をするとプライベート DMP は素晴らしいシステムに思えますが、企業側でのデータ整備の必要があるので、導入の敷居はオープン DMP よりも高いです。さらにシステム面以外にも、データの取り扱いや、各部署・担当者の連携や責任の範囲、セキュリティーやプライバシーの問題など、導入までに乗り越えなくてはならない壁も多いです。(https://dmlab.jp/words/d016.html)

ファイルに挿入され、次のページアクセスによって引き起こされる後続の RTB で使用されます。したがって、RTB は追跡の原因であると同時に追跡の手段でもあります。

5.1.1. ウェブ上の RTB：Cookie の同期

Cookie 同期は、Web トラッカーが Cookie を相互にリンクし、ある会社がユーザーに関して持っているデータを、他の会社が持つデータと結合するために利用する方法です。

機械的には非常に簡単です。ひとつのトラッキングドメインが別のトラッカーへのリクエストを引き起こします。このリクエストでは、最初のトラッカーが独自のトラッキング Cookie のコピーを送信します。2 番目のトラッカーは、独自の Cookie と最初のトラッカーからの Cookie の両方を取得します。これにより、ユーザーのプロファイルを構築しながら、他のトラッカーと「notes を比較」することができます。

Cookie 共有は、RTB の一部として一般的に使用されています。入札リクエストでは、SSP は独自の Cookie ID をすべての潜在的な入札者と共有します。同期を行わない場合、デマンドサイドプラットフォームは、独自の Cookie ID にリンクされたユーザーに関する独自のプロファイルを持つ場合があります。DSP は、Doubleclick（Google の広告ネットワーク）のユーザー「abc」が自身のユーザー「xyz」と同じであることを知らない場合があります。Cookie の同期は、これを確実なものにします。入札プロセスの一環として、SSP は通常、一度に多くの DSP への cookie-sync 要求を引き起こします。そうすることで、次回 SSP が入札リクエストを送信するときに、入札する DSP はユーザーに関する独自の行動プロファイルを利用して、入札方法を決定することができます。

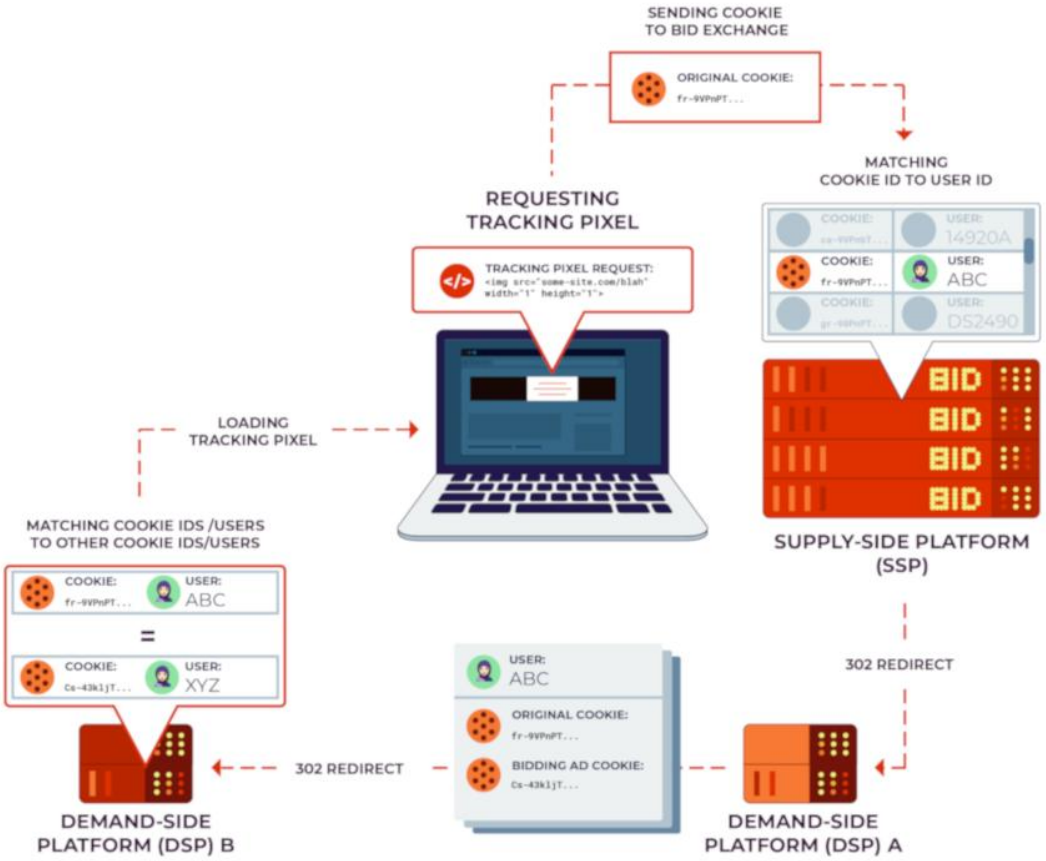

図 5.1.4: クッキーの同期。 ページ上の非表示の「ピクセル」要素は、ユーザーを DSP にリダイレクトするアドエクスチェンジまたは SSP へのリクエストをトリガーします。 リダイレクト URL には、DSP が独自の識別子にリンクさせる SSP の Cookie に関する情報が含まれています。単一の SSP が一度に多くの異なる DSP への Cookie 同期をトリガーすることがあります。

5.1.2. モバイルアプリの RTB

RTB は Web 用に作成されたものですが、モバイルアプリの広告でも同様に機能します。Cookie の代わりに、トラッカーは広告 ID を使用します。 iOS と Android に組み込まれた広告 ID により、トラッカーの仕事は容易になります。ウェブでは、各広告主が独自の Cookie ID を持ち、デマンドサイドプラットフォームは、データを特定のユーザーに関連付けるために、データを DMP と相互に同期する必要があります。

ただし、モバイルデバイスでは、各ユーザーがすべてのアプリからアクセスできる単一のユニバーサル広告 ID を持っています。つまり、上記のウェブでの同期手順はモバイルでは不要なのです。広告主は、広告 ID を使用して ID を確認し、データを共有し、入札のベースとなる詳細なプロファイルを作成することができます。

5.2. グループターゲティングと類似オーディエンス

場合によっては、大規模なプラットフォームはデータを公開しません。むしろ、データを活用したツールへの一時的なアクセスをリースします。Facebook、Google、Twitter はすべて、広告主が広告を表示させている人々のカテゴリをターゲットにできるようにします。たとえば、Facebook を使用すると、広告主は特定の「関心」や「類似性」を持つユーザーをターゲットにすることができます。

企業は、広告主に、キャンペーンの対象となる個人の実際の ID を表示しません。「サンディエゴのローラーダービー[50]に興味がある人」をターゲットにした Facebook キャンペーンを開始した場合、すぐに名前のリストを表示することはできません。ただし、この種のターゲティングを使用すると、広告主はローラーダービーに行くサンディエゴの住民に直接連絡して、外部の Web サイトまたはアプリに誘導することができます。ターゲットユーザーが広告をクリックすると、Facebook から広告主のドメインにリダイレクトされます。この時点で、広告主は彼らが Facebook からやって来たもので、ターゲットユーザー層の一部であることがわかっています。ユーザーがサードパーティのサイトにアクセスすると、広告主はデータ交換サービスを使用して、行動プロファイルや実世界のアイデンティティと一致させることができます。

さらに、Facebook により、広告主は他の人々のグループに基づいて「類似オーディエンス」[51]を構築できます。たとえば、あなたがウェブサイトを持つ消費者金融会社(payday loan company)だとします。債務者がアクセスするページに非表示の Facebook ピクセルをインストールし、そのページにアクセスしたユーザーのリストを作成してから、Facebook が「類似」していると考えるユーザーの「類似」オーディエンスの作成を Facebook の依頼することができます。次に、それらの人々を Facebook の広告でターゲティングし、Web サイトにリダイレクトします。そこで Web サイトでは、Cookie およびデータ交換を使用して、彼らが誰であるかを識別することができます。

これらの「類似」機能はブラックボックスです。それらを監査または調査する能力がなければ、どのような種類のデータを使用し、どのような種類のユーザーに関する情報を公開するのかを知ることは不可能です。広告主は、詳細情報を開示し、独立した検証を許すべきであると考えます。

50 訳注：ローラーゲームのこと。(https://ja.wikipedia.org/wiki/%E3%83%AD%E3%83%BC%E3%83%A9%E3%83%BC%E3%82%B2%E3%83%BC%E3%83%A0)

51 訳注： Look-alike audience 「Look-alike targetting」の解説として以下のものがある。「Look-Alike Targeting

Look-Alike Targeting (ルックアライ・クターゲティング) とは、広告を配信する対象を、狙っているオーディエンスと似ている相手に拡張していくことを指しています。“似ている相手”というところが特徴的です。 サイトを訪問してきた利用者や、その利用者に近い行動をするような利用者に対して広告を配信でアプローチしていくことで、広告の効果を高めていくことを期待することができます。サイトを訪れたことのあるユーザーに限定して情報や広告を配信していくよりも、より多くのユーザーに対して広告を広めていくことができるという利点があると考えられています。Look-Alike Targeting は、「オーディエンス拡張」とも呼ばれています。 」(https://www.adinte.jp/yougo/look-alike-targeting/)

5.3. データブローカー

データブローカーは、データを収集、集計、処理、および販売する会社です。それらは通常のユーザーからは見えませんが、データ共有経済の中心をなしています。多くの場合、データブローカーはユーザーとまったく直接的な関係を持たないため、データを販売している人はブローカーが存在していることに気付きません。データブローカーは、小売業者、金融テクノロジー企業、医学研究会社、オンライン広告主、携帯電話プロバイダー、IoTデバイスメーカー、地方自治体などを含むさまざまな中小企業から情報を購入します。その後、広告主、不動産業者、市場調査会社、大学、政府、民間賞金稼ぎ、その他のデータブローカーにデータまたはデータベースのサービスを販売します。

ここではカバーするには広すぎる別のトピックであり、データ販売エコシステムについて詳細に記述しているものが別にあります。企業の監視に関する Cracked Labs のレポートは、アクセス可能で詳細なものですです。World Privacy Forum の Pam Dixon は、2014 年の報告や 2015 年と 2019 年の上院前の証言など、データブローカーに関する優れたレポートも出しています。

「データブローカー」という用語は広く、電話番号または電子メールの厳選されたリストを集めて販売する零細なマーケティング会社や、数千の異なるストリームからデータを取り込み、他のビジネスにデータベースのサービスを提供する Oracle などの巨大企業まで含まれます。

一部のブローカーは、情報の生のデータの流れを販売しています。これには、小売購入行動に関するデータ、IoT デバイスからのデータ、コネクテッドカー[52]からのデータが含まれます。その他は、あらゆる種類のデータの買い手と売り手の間の情報センターとして機能します。たとえば、Narrative は売り手が「(彼らの)データの価値の機会を開き」、バイヤーが「(必要なデータ)にアクセスできるように」します。Dawex は、「あなたが直接出会い、データを直接売買することができるグローバルなデータ市場」を標榜しています。

別のクラスの企業は、仲買人または「アグリゲーター」として機能し、複数の異なるソースから生データの使用許可を得て、処理し、他のビジネスの特定のサービスとして再パッケージ化します。たとえば、主な電話会社は位置データへのアクセスを Zumigo および Microbilt と呼ばれるアグリゲーターに販売していましたが、これはさらに広範な企業へのアクセスを販売し、結果として市場は最終的に保釈金立て替え業者(bail bondsmen)や報賞金稼ぎ (および潜入リポーター) にまで手をのばしました。EFF は現在、ユーザーの同意なしにデータを販売し、そのプライバシー慣行について大衆に誤解を与えているとして、AT＆T を提訴しています。

大手データブローカーの多くは、収集した生データを販売していません。その代わりに、何千もの異なるソースからデータを収集し、これを利用して独自のプロファイルを組み立て、個人に関する推論を引き出しています。世界最大のデータブローカーの 1 つである Oracle は、ウェブ上の最大のサードパーティトラッカーの 1 つである Bluekai を所有しています。Equifax や Experian などの信用調査機関も特に活発に活動しています。米国公正信用報告法(U.S. Fair Credit Reporting Act)は、信用格付者が特定の種類のデータを共有する方法を規定していますが、取引情報や閲覧履歴など、今日トラッカーが収集するほとんどの情報を信用機関が販売することを妨げていません。これらの企業の多くは、ユーザーの行動を記述する「生得的な」特性であるサイコグラフィックス[53]を導き出す能力があると宣伝しています。たとえば、Experian は人々を「クレジットハングリーカードスイッチャー」、「規律ある従順な借入人」、「不安定な債務依存者」などの金融カテゴリに分類し、米国人口の 95%をカバーしていると主張しています。ケンブリッジアナリティカは、Facebook の「いいね」のデータを使用して悪名を馳せ、「OCEAN scores」[54] (開放性、誠実性、外向性、同意性、神経症性の評価[55]) を導くために、数百万人の有権

52 訳注：インターネットへのアクセス機能を持ち、それ自身が情報端末となって、交通情報などを含めたさまざまな情報をやりとりして運転手や乗員を支援する自動車。(英辞郎)
53 訳注：個人や住民の行動・価値観・興味などに関する研究で、マーケティングや人口動態学などで利用される。(英辞郎)
54 訳注：OCEAN は openness, conscientiousness, extraversion, agreeableness, neuroticism の頭文字。
55 訳注： OCEAN は openness、conscientiousness、extraversion、agreeableness、neuroticism の頭文字。

者情報を選挙キャンペーンに売りました。

最後に、多くのブローカーは内部プロファイルを使用して、「ID データ検証/照合」[56]または「職務充実化」サービスを他のユーザーに提供します。 Cookie やメールアドレスなどの識別子が 1 つあるビジネスは、データブローカーに支払いを行って、そのデータを「充実」させ、個人に関するその他の情報を学習させることができます。また、ある識別子（Cookie など）に関連付けられたデータを別の識別子（モバイル広告 ID など）のデータにリンクすることもできます。リアルタイム入札業者の世界では、これらのサービスは「データ管理プラットフォーム」として知られています。リアルタイム入札業者はこうした類いのサービスを利用して、入札リクエストを含む ID だけに基づいて、特定のユーザーが誰で、興味が何であるかを学習します。

何年もの間、データブローカーは目に見えず、一般の人々の心からも離れて運営されてきました。しかし、転換点に近づいているかもしれません。2018 年、バーモント州は、サードパーティとデータを売買する企業を州総長官[57]に登録することを義務付ける国内初の法律を可決しました。その結果、120 を超えるデータブローカーのリストとそのビジネスモデルに関する情報にアクセスできるようになりました。さらに、2020 年にカリフォルニア州消費者プライバシー法が施行されると、消費者はブローカーが保有する個人情報に無料でアクセスし、オプトアウトする権利を持ちます。

5.4. データ利用者

これまで、データの収集、共有、販売の方法について説明してきました。しかし、それはどこで終わるのでしょうか？個人データの消費者は誰なのでしょうか？また、それによって何をしようというのでしょうか？

5.4.1. ターゲット広告

圧倒的に大きく、最も目に見えて、どこにでもあるデータ消費者は、ターゲットを絞った広告主です。ターゲットを絞った広告によって、広告主はユーザーの属性、サイコグラフィックス、その他の特性に基づいてユーザーに到達することができます。行動ターゲティング広告は、パーソナライズされた広告を作成するために、ユーザーの過去の行動に関するデータを活用するターゲット広告のサブセット[58]です。

最大のデータコレクターは、最大のターゲット広告主でもあります。 Google と Facebook はともに、米国のデジタル広告市場のほぼ 60％を管理しており、そのためにそれぞれのデータを使用しています。Google、Facebook、Amazon、Twitter は端末相互間のターゲティングサービスを提供しており、広告主は高レベルのカテゴリのユーザーをターゲティングでき、広告主自身がデータにアクセスする必要はありません。Facebook では、広告主はロケーション――つまり、年齢、性別、教育、収入などの人口統計、趣味、音楽のジャンル、有名人、政治的偏見などの関心事――に基づいてユーザーをターゲットにできます。Facebook が使用する「関心」の一部は、ユーザーが「いいね」またはコメントした内容に基づいており、その他は Facebook のサードパーティ追跡に基づいて導出されて

56 訳注：Identity Resolution。LiveRamp という会社は「Identity Resolution(ID データ検証/照合)により顧客レベルでのマーケティングを可能にします。あらゆるチャネルにおいて cookie でなく、人ベースでのマーケティングが可能です。」「Identity Resolution を用いることで、パーソナライズ化されたコミュニケーションによるターゲティング、キャンペーンの効果測定をオムニチャネルの観点から行うことができます。」などの紹介がある。(https://liveramp.co.jp/identity-resolution/) また、インフォマティカという会社は、「民間企業や政府機関の ID データ検索／照合に使える、非常に拡張性の高いソフトウェア」「隠れた関係性を特定することで、不正行為を検出し、リスクを回避し、コンプライアンスを徹底できます」として、テロ資金、マネーロンダリング対策などのリスク管理に有効と述べている。(https://www.informatica.com/jp/products/master-data-management/identity-resolution.html)

57 訳注：文書の記録や法令の発布、選挙管理などを行う。(英辞郎)

58 訳注：サブセット 部分集合のことだが、IT の分野では、仕様や規格、ソフトウェアなどの機能についてよく用いられる用語で、本来備えている要素の一部を省略・削除した簡易版・限定版という意味で用いられることが多い。(http://e-words.jp/w/%E3%82%B5%E3%83%96%E3%82%BB%E3%83%83%E3%83%88.html)ここでは、行動ターゲティング広告がターゲット広告の一部の機能であるという意味。

います。Facebookはデータを利用して広告主を対象ユーザーとマッチングさせるのにこのデータを利用する一方で、Facebookはこれらの広告主とこのデータを共有しません。

リアルタイム入札業者（RTB）[59]はより多くのデータ共有を必要とし、さまざまなレベルのプロセスに関与している膨大な数の中小企業が存在します。大手ハイテク企業もこの分野でサービスを提供しています。GoogleのDoubleclick Bid ManagerとAmazon DSPはどちらもRTBデマンドサイドプラットフォームです。RTBでは、広告主自身（またはエージェント）が各個人に到達するかどうか、どの広告を表示するかを決定できるように、識別子が共有されます。RTBエコシステムでは、広告主はユーザーの行動に関する独自のデータを収集し、社内の機械学習モデルを使用して、どのユーザーが広告に関与するか、製品を購入する可能性が最も高いかを予測します。

一部の広告主は、FacebookまたはGoogleのユーザーに到達したいのだが、大企業独自のターゲティング技術は使用したいとは考えていません。代わりに、データブローカーから連絡先情報のリストを購入し、これらのリストをFacebookまたはGoogleに直接アップロードすることで、すべてのプラットフォームのユーザーに到達することができます。このシステムは、差別的または悪意のあるターゲティングを抑制しようという大企業の努力を損うことになります。GoogleやFacebookのようなターゲティングプラットフォームでは、広告主が特定のエスニシテイィのユーザーを仕事、住宅、クレジットの広告でターゲティングすることはできません。ただし、広告主はデータブローカーから個人に関する人口統計情報を購入し、同じ人種グループに属する名前のリストをアップロードし、プラットフォームでそれらの人々を直接ターゲットにすることができます。GoogleとFacebookは、連絡先リストを持つユーザーをターゲットにするために「機密情報」を使用することを禁止していますが、これらのポリシーをどのように実現しているのかは不明です。

5.4.2. 政治キャンペーンと利益団体

データ収集とターゲット広告から利益を得ようとするのは企業だけではありません。Cambridge Analyticaは、不正な個人データを使用して数百万人の有権者の「サイコグラフィックス[60]」を推定し、そのデータを使用して政治キャンペーンを支援しました。2018年、グループCatholicVoteは携帯電話の位置データを使用して、誰がカトリック教会の中にいたのかを判断し、「投票推進」広告で彼らをターゲットにしました。反中絶グループも、同様のジオフェンシング・テクノロジーを使用して、女性が妊娠中絶クリニックにいる時にこうした女性を広告のターゲットに設定しました。

こうしたことは、個々ばらばらなことではありません。寄付に依存する一部の非営利団体は、潜在的な寄付者を絞り込むのに役立つデータを購入します。全国の多くの政治家は、有権者をターゲットにするために公開の有権者登録データを使用しています。民主党全国委員会は、2020年の選挙に先立ち、「データウェアハウス」に多額の投資を行っていると伝えられています。そして共和党のコンサルティング会社であるDeep Root Analyticsは、米国有権者データの歴史上最大の侵害の原因を引き起しました。ほぼ2億人のアメリカ人に関する名前、登録の詳細、および「モデル化された」エスニシティと宗教のデータを収集していたのです。

5.4.3. 借金取り、賞金稼ぎ、詐欺捜査官

借金取り、賞金稼ぎ、差し押さえ機関は皆多くの情報源からロケーションデータを購入して利用しています。EFFは、位置データを情報収集サイト(アグリゲーター)に販売する役

59 訳注：RTB ネット広告の取引手法の一つで、広告が表示される瞬間に複数の広告代理店や広告主で入札を行い、最も条件の良い広告を掲載する方式。広告主は掲載したい広告と希望掲載枠、掲載条件、希望落札価格などを入力し、入札を待つ。広告掲載メディアに閲覧者がアクセスし、広告枠に広告が表示される瞬間、複数の広告主の落札条件を突き合わせ、最も高い価格を提示していた広告主の広告が掲載される。この処理は1秒に満たないわずかな時間で実行され、メディアや広告の閲覧に支障をきたすことはない。
　掲載枠が表示されるたびにこのような落札を実施し、その都度、落札者と価格が決定されてゆく。広告主と掲載メディアの双方にとって、常に広告市場の実勢に応じた価格・条件での取引が可能となる。(http://e-words.jp/w/RTB.html)

60 訳注：個人や住民の行動・価値観・興味などに関する研究(英辞郎)

割を担ったことに対して AT&T を訴えています。AT&T のこうした振舞いによって、奨金稼ぎがアクセスできる流通市場が実現してしまったのです。ただし、データの情報源は携帯電話会社だけではありません。保釈金立替会社 Captira は、携帯電話と ALPR から収集した位置データをわずか 7.5 ドルで賞金稼ぎに販売しました。そして、数千のアプリが GPS の利用許可を用いて「同意のもとに」位置データを収集し、そのデータをダウンストリーム[61]アグリゲーターに販売します。このデータは、逃亡者、債務者、および自動車代金未払い者を見つけるために利用できます。また、調査によって明かなように、ほぼ誰でも購入して悪用することができます。

5.4.4. 都市、法執行機関、intelligence 報機関

また、公共部門は、あらゆる種類のアプリケーションのために民間部門からデータを購入します。たとえば、米国移民税関当局は、政府機関が 1 国外追放を予定している人々を特定するために、Vigilant[62]から ALPR データを購入しました。政府機関は、カリフォルニア州都市連盟(California Cities)によると、政府機関とデータブローカー間の契約に関する州の消費者データプライバシー法の例外を求める書簡において、人的サービスの適格性の決定から徴税まで無数の業務でデータブローカーと契約しています。このことを問題にしてきた人たちは、政府機関と民間データブローカーの間のこうした取り決めが、消費者のデータプライバシーに対する脅威であり、同時に、政府自身のデータベースの法的規制への回避策だとして非難しています。そしてもちろん、国家安全保障による監視は、多くの場合、民間企業の大量に蓄積された消費者データのデータマイニング[63]に基づいています。たとえば、エドワード・スノーデンが明らかにした PRISM プログラムの一環として、NSA は Google、YouTube、Facebook、Yahoo から直接個人データを収集していました。

6. 反撃

侵襲的または操作的な広告のターゲットにならないように、トラッキングに抵抗することもできます。あなたの個人情報があなたの背後で交換され売られていることに不満を感じるかもしれません。あなたに危害を加えたいと願う誰かが、サードパーティのデータブローカーを通してあなたの居場所にアクセスできることを心配するかもしれません。企業によって収集されたデータが警察や諜報機関の手に渡ることを、多分、恐れるかもしれません。または、サードパーティのトラッキングは、漠然とした不安感を与えながらずっと続いていく迷惑行為にすぎない場合もあります。

しかし、残念なことに、トラッキングを避けるのは困難なことなのです。何千もの独立したアクターが何百もの異なる技術を使用しているため、企業の監視はくまなく行き渡り、資金も豊富です。すべてのトラッキングの手段を阻止しうるようなどんでん返しはありませんが、プライバシーを取り戻すためにできることはまだたくさんあります。このセクションでは、プライバシーを重視するユーザーがサードパーティのトラッキングを回避あるいは混乱させるいくつかの方法について説明します。

個人は、プライバシーを保護するためにどれだけの労力を費やすかを自分で決める必要があります。EFF のトラッカーブロッカー拡張機能の Privacy Badger をブラウザにインストールしたり、電話の設定を変更したりするなど、小さな変更により、トラッカーが収集して共有できるデータの量を大幅に削減できます。サードパーティアプリのアンインストールや Tor の使用などの大きな変更は、時間、利便性、そして時にはお金を犠牲にして、より強力なプライバシー保証を提供することができます。深刻な懸念をもっているユー

61 訳注：ダウンストリーム 通信回線で、ネットワークの中心側（通信施設や集線装置、サーバなど）から末端側（個々の利用者や端末、クライアントなど）へ向かう方向。また、その方向に流れるデータや信号、およびその通信速度などのこと。(http://e-words.jp/w/%E3%83%80%E3%82%A6%E3%83%B3%E3%82%B9%E3%83%88%E3%83%AA%E3%83%BC%E3%83%A0.html)

62 訳注： 米国のデータ会社を指すと思われる。https://www.vigilantsolutions.com/ EFF Responds to Vigilant Solutions' Accusations About EFF ALPR Report EFF Calls on California to End Vendor-Driven ALPR Training

63 訳注：データマイニング 情報システムに蓄積した巨大なデータの集合をコンピュータによって解析し、これまで知られていなかった規則性や傾向など、何らかの有用な知見を得ること。(http://e-words.jp/w/%E3%83%87%E3%83%BC%E3%82%BF%E3%83%9E%E3%82%A4%E3%83%8B%E3%83%B3%E3%82%B0.html)

ザーにとっては、より強力な手段をとることに価値があります。

最後に、これはあなたの責任ではないことに注意してください。プライバシーは個人の責任の問題であってはなりません。あなたをひそかに監視できる最新のテクノロジーに忙殺されるのはあなたの仕事ではありません。また、電話がどのようにしてデータを共有するのかを理解するためにプライバシーポリシーの 25 万語もの法律用語を読む必要もありません。プライバシーは権利なのであり、教育水準の高い人や暇な人の特権ではありません。誰もが自分のプライバシーを尊重するオンラインとオフラインの世界に住むのに値する人間なのです。

より良い世界では、私たち自らが自分のデータを共有することを選択した企業が、私たちの信頼を獲得し、他の誰も私事に介入しないようになるでしょう。これが、EFF が企業に消費者のデータプライバシーを尊重するように訴える訴訟を起こしている理由であり、プライバシーが国の法となるような立法を支持する理由なのです。EFF のメンバーとサポーターの助けを借りて前進していますが、企業監視ポリシーを変えるには紆余曲折があります。それでは、今のところ、どのように反撃できるかについて話しましょう。

6.1. web 上で

Web 上のトラッキングに晒されないように制限をかけるいくつかの方法があります。まず、ブラウザの選択が重要です。特定のブラウザ開発者は、あなたにとって利益になるように行動する「ユーザーエージェント」としてのソフトウェアの役割をより一層重視しています。 Apple の Safari は、サードパーティ Cookie、ファーストパーティとサードパーティ

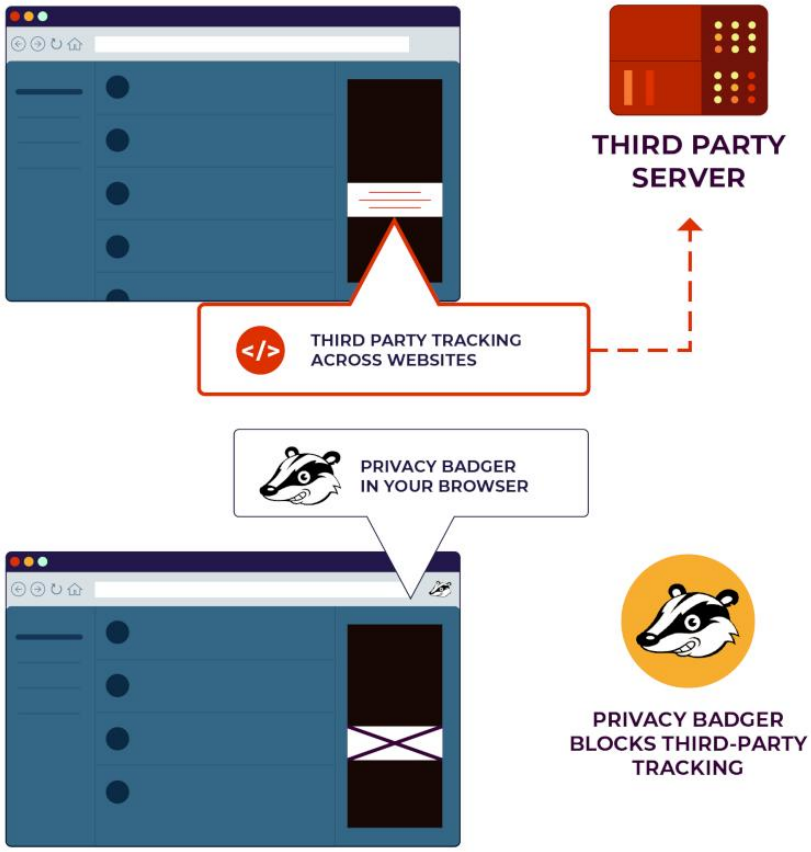

図 6.1.1: EFF の Privacy Badger などのブラウザ拡張機能は、ウェブ上のサードパーティのトラッキングに対するプライバシー保護のレイヤーを提供します。プライバシーバジャーは、ヒューリスティックを使用してトラッカーをブロックすることを学習します。つまり、静的なリストベースのブロッカーが見逃している新規または一般的でないトラッカーをキャッチする可能性があります。

の Cookie 共有、フィンガープリントなど、最も一般的なトラッキング形式に対して積極的な対策を講じています。Mozilla の Firefox は、デフォルトで既知のトラッカーからのサードパーティ Cookie をブロックし、Firefox のプライベートブラウジングモードはトラッカーへのリクエストを完全にブロックします。

EFF の Privacy Badger や uBlock Origin などのブラウザー拡張機能は、プライバシー保護の別のレイヤー[64]を提供します。特に、Privacy Badger は、ヒューリスティック[65]を使用してトラッカーをブロックすることを学習します。つまり、静的なリストベースのブロッカーが見逃している新規または一般的でないトラッカーをキャッチする可能性があります。これにより、Privacy Badger は、切断リストに依存する Firefox が提供する組み込みのプライバシー保護の優れた補足機能となります。また、Google Chrome はデフォルトではトラッキング動作をブロックしませんが、Chrome にプライバシーバジャーまたは別のトラッカーブロッキング拡張機能をインストールすると、比較的プライバシーの露出を少なくして利用できます。（ただし、Chrome で計画中の変更では、多くのユーザーがトラッキングをブロックするために利用するセキュリティやプライバシーツールに影響を与える可能性があります。）

完璧なトラッカーブロッカーはありません。すべてのトラッカーブロッカーでは、正当なコンテンツを提供する企業に対する例外を設定をする必要があります。たとえば、Privacy Badger は、トラッキング動作を実行することが知られているドメインのリストを維持し、コンテンツ配信ネットワークやビデオホストなど、多くのサイトが機能するのに必要なコンテンツを提供します。Privay Badger は、Cookie とローカルストレージへのアクセスをブロックすることにより、これらのドメインのトラッキング機能を制限しますが、専用のトラッカーは引き続き IP アドレス、TLS 状態、およびある種のフィンガープリント可能なデータにアクセスすることができます。

さらに一歩進んで、あれこれやってみたい場合は、自宅にネットワークレベルのフィルターをインストールできます。Pi-hole は、ローカルネットワーク上のすべてのトラフィックを DNS[66]レベルでフィルタリングします。これは個人 DNS サーバーとして機能し、トラッカーをホストすることが知られているドメインへのリクエストを拒否します。 Pi-hole は、スマート TV、ゲームコンソール、モノのインターネット製品など、他の方法では構成が困難なデバイスからのリクエストの追跡をブロックします。

露出を可能な限り減らしたい人にとっては、Tor Browser はプライバシーの金字塔ともいえるものです。Tor は、オニオンルーティングサービスを使用して、ユーザーの IP アドレスを完全に隠しします。デフォルトで HTMLcanvas[67]へのアクセスをブロックするなど、フィンガープリントを減らす積極的な手順をとっています。Tor は、TLS セッションチケットを完全に拒否し、各セッションの終了時に Cookie を消去します。

残念ながら、2019 年において、Tor でウェブを閲覧するのは万人向けではありません。トラフィックが大幅に遅くなるため、ページの読み込みに非常に時間がかかり、ビデオその他のリアルタイムコンテンツのストリーミングは極めて困難です。さらに悪いことに、現

64 訳注：コンピュータなどの通信機器は、いくつかの階層からなる構造をもっている。基本ソフト(OS)が属する「層」、ワープロソフトなどユーザが仕事に使うソフトウェアが属する「層」、ネットワークの通信を行なうための「層」などが重なりあう。こうした「層」をレイヤと呼ぶ。

65 訳注： ヒューリスティック 「問題の解答を得るための方法論の一つで、常に正しいとは限らないが経験的にある程度正しい解を導ける推論や経験則などを利用して、近似的あるいは暫定的な解を得る手法のことをヒューリスティックということが多い。
　コンピュータプログラムのアルゴリズム（計算手順）などで用いられる概念で、正しい解を厳密に求めるには複雑な手順や膨大な計算が必要な場合や、既知の情報や手順では正しい解を得られない場合（新種のコンピュータウイルスの検知など）などで利用される。広義には近似アルゴリズムの一種だが、理論的に解の精度が保証されるものを（狭義の）近似アルゴリズム、精度が保証されないものをヒューリスティックとして区別することがある。」(http://e-words.jp/w/%E3%83%92%E3%83%A5%E3%83%BC%E3%83%AA%E3%82%B9%E3%83%86%E3%82%A3%E3%83%83%E3%82%AF.html)

66 訳注： DNS インターネットなどの TCP/IP ネットワーク上でドメイン名やホスト名と IP アドレスの対応関係を管理するシステム。(http://e-words.jp/w/DNS.html)

67 訳注： HTML はウエッブを作成するための言語。canvas は、HTML 言語でグラフィックを扱うためのの仕組み。

代の Web の多くは、「疑わしい」と見なされるソースからのトラフィックをブロックまたはスロットルする invisible CAPTCHA[68]に依存しています。Tor からのトラフィックは、高リスクとして分類されることが多いため、Tor を使用した Google 検索のような簡単な操作で CAPTCHA テストが引き起こされることがあります。また、Tor は攻撃者も使用するパブリックネットワークであるため、一部の Web サイトは Tor の訪問者を完全にブロックしています。

6.1.1. 携帯電話で

モバイルデバイスでのトラッカーのブロックは更に複雑です。ブラウザや拡張機能など、多くの基盤をカバーできる解決策はありません。残念ながら、特定のデバイスで特定の種類の追跡を制御することは不可能なのです。

追跡に対する最初の防衛線は、デバイスの設定です。

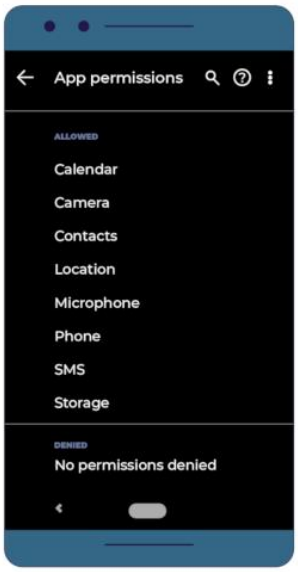

図 6.1.2: アプリの許可ページ。 "width =" 1081 "height =" 1849 ">

iOS と Android は、ユーザーによって各アプリのアクセス許可を表示および制御することができます。アプリに与えられている権限を確認し、不要な権限を削除する必要があります。また使用していないアプリを削除することです。アプリごとの設定に加えて、デバイスがロケーションなどの特に機密性の高い情報を収集・共有する方法に影響を与えるグローバル設定も変更することができます。また、アプリが使用されていないときにインターネットへのアクセスを許可する方法も制御できます。こうしたことにより、パッシブトラッキングを防ぐことができます。

iOS と Android のオペレーティングシステムには、さまざまな方法でデバイスの広告 ID をリセットするオプションもあります。iOS では、ゼロの文字列に設定することにより、広告 ID を完全に削除できます。 （ここに iOS で広告追跡をブロックする別の方法があります。）Android では、手動でリセットできます。これは、Cookie をクリアすることと同じですが、新しい Cookie をブロックするわけではありません。追跡を完全に無効にするわけではありませんが、トラッカーにとっては、あなたに関する統一されたプロファイルを作成するのが難しくなります。

Android には、「インタレストベース広告をオプトアウト」する設定もあります。これにより、ユーザーがターゲット広告に自分のデータを利用されたくないという信号をアプリに送信しますが、実際にはアプリがそうすることを止めさせるわけではありません。実際、最近の調査では、何万ものアプリが単にこの信号を無視しているだけだということがわかっています。

iOS には、他のアプリからのトラッキング行動をフィルタリングできるアプリがいくつかあります。Android では、それほど簡単ではありません。Google は、アプリストアの Play ストアの広告ブロッカーやトラッカーブロッカーを禁止しているため、この種のもので公式のアプリはありません。Play ストア以外でブロッカーを「サイドロード」することは可能ですが、これは非常に危険です。できればオープンソース[69]を使用していて、信頼できる発行元からのアプリのみをインストールしてください。

68 訳注： invisible CAPCHA サイトにアクセスするときに、人間なのかロボットなのかを識別するために、画像や文字を選択させるなどの仕組みが組み込まれていることがある。こうした手間を省いて、ログインと同時にセキュリティチェックが行なえるような仕組みが invisible CAPTHA。

69 訳注： オープンソース 人間が理解しやすいプログラミング言語で書かれたコンピュータプログラムであるソースコードを広く一般に公開し、誰でも自由に扱ってよいとする考え方。また、そのような考えに基づいて公開されたソフトウェアのこと。(http://e-words.jp/w/%E3%82%AA%E3%83%BC%E3%83%97%E3%83%B3%E3%82%BD%E3%83%BC%E3%82%B9.html)

また、デバイスが通信しているネットワークについても考慮する必要があります。なじみのない公衆 WiFi ネットワークへの接続を避けるのが最善です。おそらく「無料」の WiFi の利用は、あなたのデータ漏洩のリスクという対価を払うことになるかもしれません。

ワイヤレスビーコンもデバイスから情報を収集しようとしています。デバイスがハードウェア MAC アドレスをブロードキャストしている場合にのみ、識別情報を収集できます。iOS と Android の両方がデフォルトでこれらの MAC アドレスをランダム化しますが、他の種類のデバイスはそうではなくて、電子リーダー、スマートウォッチ、または車の場合には、プローブ要求[70]をブロードキャストしている可能性があります。プローブ要求は、位置データの取得に使用できます。これを防ぐには、通常、WiFi と Bluetooth をオフにするか、デバイスを「機内モード」に設定します（これはバッテリーを節約する良い方法です！）。

最後に、本当に匿名にする必要がある場合、「プリペイド携帯電話」[71](burner phone)を使用すると、固有のハードウェア識別子に関連付けられた追跡を制御するのに役立ちます。

6.1.2. IRL

現実の世界では、オプトアウトはそれほど単純なことではありません。

既に説明したように、デバイスが作動しないように変更するための多くの方法があります。しかし、顔認識カメラと自動ナンバープレートリーダーによるトラッキングを避けることはほとんど不可能です。もちろん、顔認識アルゴリズムを混乱させるために顔をペイントしたり、ALPR 企業のデータベースを回避するために車を持たない選択をしたり、現金または仮想クレジットカードを使用して支払い処理者のプロファイリングを止めたりすることはできます。しかし、ほとんどの場合、これらのオプションは大半の人にとっては現実的ではなく、誰もが直面しているあらゆるトラッキングを回避することは不可能です。

しかし、知識は戦いの半面にすぎません。今のところ、顔認識カメラは、海外旅行中に空港などの特定の場所であなたを識別する可能性が最も高いものです。ナンバープレート読み取り監視カメラははるかに普及しており、回避するのが困難ですが、どうしても必要な場合は、公共交通機関やその他の交通手段を使用して、車両の追跡頻度を制限することができます。

6.1.3. 議会

一部の地域では、ユーザーをトラッキングから保護する法律があります。欧州連合の一般データ保護規則（GDPR）は、収集された情報にアクセスしてこれを削除する権利を与えています。また、企業は、データを利用する正当な理由を必要とします。これには、「正当な利益」またはオプトイン同意[72]が含まれます。GDPR は完璧からははほど遠いもので、その有効性は、規制当局と裁判所が今後どのように実施するかにかかっています。しかし、ユーザーに意味ある権利を与え、これに違反する企業に実質的な影響をもたらします。

70 訳注： プローブ要求 クライアントはビーコンを受信すると、自身が設定している ESS-ID(無線ネットワーク名のこと)かどうかをアクセスポイントに対して問い合わせます。この問い合わせを「Probe Request（プローブ要求）」と呼びます。もし、同じ ESS-ID であれば、アクセスポイントは返事を返します。この返事を「Probe Response（プローブ応答）」と呼びます。(https://www.itbook.info/network/wlc1.html)

71 訳注： 日本では、ソフトバンクが 2019 年 11 月にサービスを終了し、au も 2021 年にサービスを終了する予告が出されている。NTT はすでに 2005 年にサービスを停止している。いずれも「オレオレ詐欺」など犯罪に利用されることを理由にサービスを停止する方向にある。

72 訳注： オプトイン opt-in 加入や参加、許諾、承認などの意思を相手方に明示すること。個人が企業などに対し、メールなどのメッセージの送信や、個人情報の収集や利用などを承諾する手続きなどを指すことが多い。
　個人情報の利用などに関して、対象者から明確に許諾を得ない限り実施しない（あるいは、してはならない）とする原則のことを「オプトイン方式」ということがある。
　これに対し、離脱や脱退、拒否、停止、中止などの意思を表明したり申し入れることを「オプトアウト」（opt-out）という。(http://e-words.jp/w/%E3%82%AA%E3%83%97%E3%83%88%E3%82%A4%E3%83%B3.html)

米国では、州法および連邦法のごく一部が特定のプライバシー保護を提供しています。バーモント州のデータプライバシー法は、データブローカーに透明性を求めています。イリノイ州生体認証情報保護法（BIPA）では、企業は生体認証識別子を収集または共有する前にユーザーから同意を得る必要があるとしています。2020年には、カリフォルニア消費者プライバシー法（CCPA）が発効し、ユーザーは自分の個人情報にアクセスし、削除、販売をオプトアウトする権利が与えられています。一部のコミュニティは、政府による顔認証の使用を制限する法律を可決するか、近々可決する予定のところがあります。

連邦レベルでは、状況によっては、HIPAA、FERPA、COPPA、ビデオプライバシー保護法、いくつかの金融データプライバシー法などの法律によって一部の情報が保護されています。ただし、これらの特定分野に固有の連邦法は、特定の企業が保有する特定のタイプの人々に関する特定のタイプの情報にのみ適用されます。これらは、トラッカー、広告主、データブローカーに悪用されており、法と現実の間には多くのギャップがあります。

早い話、米国のほとんどのサードパーティのデータ収集は規制されていないのです。そのため、EFFはユーザーのプライバシーを保護するための新しい法律を提唱しています。人々は、自分についてどのような個人情報が収集され、何が行われているのかを知る権利を持つべきです。インフォームドオプトインの同意を与えない限り、データの企業処理はなされるべきではありません。ユーザーがプライバシー権を行使することを選択した場合、企業は追加料金を請求したり、サービスを低下させたりすすべきではありません。データを誤用したり、誤って処理したりした場合は、説明責任を負う必要があります。そして、プライバシーが侵害された場合、人々は企業を裁判にかける権利を持つべきです。

最初のステップは、マジックミラーを壊すことです。私たちは、鏡の後ろに潜むトラッカーの錯綜したネットワークを白日のもとに晒す必要があります。白日の下で、これらの商業的監視システムが何なのかが暴露されます。この監視システムは、オーウェル的ではありますが、全知ではありません。普及してはいますが、これを避けることができないわけではありません。私たちユーザーは、自分が何に直面しているのかを理解すれば、これに反撃することもできるはずです。

著者 bennett

英語版PDF

Downloads(英語版) PDF icon Behind the One-Way Mirror: A Deep Dive Into the Technology of Corporate Surveillance