1W-08

Browser Fingerprintingにおけるスマートフォンの識別に関する考察

高橋 和司† 高須 航‡ 山田 智隆‡ 武居 直樹‡ 西倉 裕太‡
石川 貴之‡ 細井 理央‡ 安田 昂樹† 平 俊介† 齋藤 孝道†

明治大学† 明治大学大学院‡

1 はじめに

端末から採取可能な複数の情報の組み合わせにより端末を識別する Browser Fingerprinting（以降，Fingerprinting という）と呼ばれる手法が Web 広告事業者を中心に利用されている．先行研究[1]では，スマートフォンにおけるインストール済みフォントリストやプラグインリストは値の種類が少なく，スマートフォン端末は PC 端末に比べて識別が困難とされている．しかし，HTML5 の API に代表されるような Web 技術の発展に伴い，採取可能な情報が増加し，スマートフォンでの識別の実現可能性が指摘された[2]．本論文では，新たに採取可能となった情報も含め，実験で採取したデータを iOS 端末と Android 端末に分けて分析し，識別可能性についてそれぞれ示す．

2 Browser Fingerprinting

2.1 定義

Web サーバがブラウザを通して採取可能なブラウザや端末の情報を特徴点という．特徴点の値を 1 つ以上組み合わせたものを Browser Fingerprint（以降，Fingerprint という）という．

2.2 スマートフォンにおける特徴点

スマートフォン特有の値をとる特徴点には，ユーザエージェント文字列（以降，UA 文字列という），グローバル IP アドレス，画面解像度およびデバイスピクセル比がある．

iOS 端末上や Android 端末上のブラウザでは，UA 文字列に iPhone や Android という文字列が含まれる．よって，UA 文字列が偽装されていない限り，スマートフォン経由のアクセスかどうかを判別することができる．特に，Android 端末においては UA 文字列中に機種名を含むので，端末を識別できる可能性が高い．

スマートフォンにおいて，各キャリアが公開している IP アドレス帯の情報や IP アドレスの検索機能を用いることで，端末のキャリアを推定できる．今回は IP アドレスからキャリアを推定する手段として，whois コマンドを用いた．ただし，Wi-Fi 環境で接続している場合や MVNO 利用である場合，キャリアは推定できないが，接続している組織名や ISP 名は推定できる．モバイル利用時，グローバル IP アドレスの値は短期間で変化し，特に LTE や 3G で接続の場合は頻繁に変化する．よって，IP アドレスの値を特徴点として使用した場合，識別能力が低下する[3]．そこで，グローバル IP アドレスを基に推定した各キャリアまたは組織名（以降，キャリア情報という）を特徴点として使用することで，時間経過による変化の少ない特徴点とした．

画面解像度からは端末の画面サイズが分かるので，端末を識別する情報の 1 つとして利用できる．

デバイスピクセル比は，ほとんどの PC 端末において値が 1 であるが，スマートフォンでは 1.5 や 3 などの多様な値をとる．iOS 端末ではデバイスピクセル比の値は 2 と 3 が確認でき，Android 端末では 7 種類の値が確認できた．

2.3 Fingerprint 採取サイト

我々の研究グループでは，Fingerprint を採取する Web サイトを運営している[4]．この Web サイトでは，同一の端末上のブラウザからのアクセスであることを確認するために，ブラウザ識別用の HTTP クッキー（以降，UID という）を生成し利用している．

3 採取データの分析

3.1 データセット

我々が運営する Web サイトで 2013 年 12 月 30 日から 2015 年 10 月 12 日の期間に採取したデータの中で，UA 文字列に iPhone または Android を含むデータを今回使用するデータセットとした．サンプル数は iOS 端末が 519 件，Android 端末が 326 件であった．UID 数は iOS 端末が 218 件，Android 端末が 162 件であった．複数回アクセスがあった UID 数は iOS 端末が 114 件，Android 端末が 75 件であった．UID は利用者により削除できるので，同一端末上のブラウザからのアクセスでも異なる UID が割り振られているデータもあることに注意されたい．

3.2 エントロピー

各特徴点がとる値の種類と値の偏りを一元的に示すため Shannon エントロピーを用いた．これを，識別能力の評価のための指標とする．

今回スマートフォンを識別するために用いた特徴

Study on Identifying Smartphone by Using Browser Fingerprinting

†Kazushi TAKAHASHI ‡Ko TAKASU
‡Tomotaka YAMADA ‡Naoki TAKEI ‡Yuta NISHIKURA
‡Takayuki ISHIKAWA ‡Rio HOSOI †Koki YASUDA
†Shunsuke TAIRA †Takamichi SAITO

†Meiji University
‡Graduate School of Meiji University

点は，2.2 節で述べた 4 つの特徴点とタイムゾーンである．エントロピーを表 1 にそれぞれ示す．また，これら個別の Fingerprint を組み合わせた Fingerprint を 5_Fingerprint とする．さらに，5_Fingerprint からデバイスピクセル比およびタイムゾーンを除いた組み合わせを 3_Fingerprint とする．除いた理由は後述する．id は本データセットにおけるエントロピーの最大値を示す．

表 1　今回使用した特徴点のエントロピー

番号	特徴点	iOS	Android
1	id	9.019591	8.348728
2	UA 文字列	5.091391	6.827818
3	キャリア情報	2.907443	2.805793
4	画面解像度	1.489445	2.72618
5	デバイスピクセル比	0.493548	1.677402
6	タイムゾーン	0.237683	0.415844
7	5_Fingerprint (#2, #3, #4, #5, #6)	7.086376	7.0219
8	3_Fingerprint (#2, #3, #4)	7.086376	7.0219

UA 文字列のエントロピーに対する 3_Fingerprint のエントロピーに着目すると，iOS 端末は Android 端末に比べて大きくなっている．これは，画面解像度が機種情報に対応しているからである．2.2 節で述べた通り Android 端末は UA 文字列に機種名を含むのでエントロピーの増加が小さく，iOS 端末は機種名を含まないので増加が大きくなったと考えられる．

デバイスピクセル比は 2.2 節で述べた通り多様な値を持つ．しかし，デバイスピクセル比の情報は機種を特定する 1 要素にとどまり，UA 文字列，キャリア情報および画面解像度から得られる情報に包含されているので，エントロピーが増加しない．

タイムゾーンのエントロピーが低かった理由として，今回アクセスがあった端末の大多数が国内からのアクセスであったことが挙げられる．海外からのアクセスが増えた場合，エントロピーが上昇することが考えられる．

3.3　スマートフォンにおける識別精度

国内からのアクセスが大多数であったデータを用いた今回の分析では，iOS，Android 端末ともに，UA 文字列，キャリア情報および画面解像度を組み合わせた情報が最もエントロピーが高まると判明した．よって，識別精度の評価には 3_Fingerprint を用いる．今回の実験では，UID 数を端末数とし，3_Fingerprint が完全一致した場合のみ同一とみなす．

本論文では，同一の端末上のブラウザからのアクセスを同一と判定できる割合（以降，TP 率という）と異なる端末上のブラウザからのアクセスを異なると判定できる割合（以降，TN 率という）を用いる．TP 率は (1) 式と，TN 率は (2) 式と定める．

$$\text{TP 率}=\frac{\text{同一の UID 内で Fingerprint が不変の UID 数}}{\text{複数回アクセスがあった UID 数}}\times 100 \quad (1)$$

$$\text{TN 率}=\frac{\text{他の UID の Fingerprint と重複しない UID 数}}{\text{全ての UID 数}}\times 100 \quad (2)$$

今回，3_Fingerprint で TP・TN 率を算出する実験（以降，実験 1 という）および 3_Fingerprint からキャリア情報を除いた組み合わせで TP・TN 率を算出する実験（以降，実験 2 という）の 2 つの実験を行った．その結果を表 2 に示す．

表 2　識別精度

		iOS	Android
実験 1	TP 率	66.67%	76.00%
	TN 率	54.59%	72.84%
実験 2	TP 率	75.44%	81.33%
	TN 率	31.19%	70.37%

表 2 より，iOS，Android 端末ともに実験 2 は実験 1 に比べ TP 率が高く，TN 率は低い．これは Wi-Fi と LTE など複数の環境からアクセスした際の変化を排除できたことで TP 率が上がり，一方でキャリア情報のみで識別できていたブラウザが存在していたことから TN 率が低下したと考えられる．

今回の実験により，TN 率は iOS 端末で 54.59%，Android 端末で 72.84%となった．一方，TP 率は iOS 端末で 66.67%，Android 端末で 76.00%となった．

4　まとめ

本論文では，iOS 端末および Android 端末における Fingerprint での識別について示した．UA 文字列，キャリア情報および画面解像度を用いることで Android 端末では高い精度で識別できたが，iOS 端末は Android 端末に比べて Fingerprint を用いた識別能力が低いことが分かった．

5　参考文献

[1] P Eckersley, How Unique Is Your Web Browser?, in Proc. of Privacy Enhancing Technologies Symposium (2010), 2010.

[2] 齋藤孝道，高須航，山田智隆，武居直樹，石川貴之，細井理央，安田昂樹，高橋和司，“Web Browser Fingerprint 技術の現状と課題”，コンピュータセキュリティシンポジウム 2015 , 2015

[3] 磯侑斗，桐生直輝，塚本耕司，高須航，山田智隆，武居直樹，齋藤孝道，“Web Browser Fingerprint を採取する Web サイトの構築と採取データの分析”，コンピュータセキュリティシンポジウム 2014 , 2014

[4] https://www.saitolab.org/fingerprint/