

Residential IP Proxy サービスを用いた位置情報・ターゲット広告の調査

福田 ひかり †井窪 竜矢 †菊池浩明 †

明治大学総合数理学部 †



1 はじめに

近年、セキュリティ対策の一環としてプロキシサービスが広く使われており、そのサービスの一部である住宅用ネットワークを利用したトラフィック中継を提供する Residential IP Proxy (以下 RESIP とする) が普及し始めた。RESIP は中継するホストを意思表示が確認できた一般住宅から収集することで成り立っている。

しかし、Mi らの調査 [1] によると、収集する際に本人の同意を得ていない場合があることが判明した。そこで本実験では、意図せず中継するホストになっていないか確認するため、RESIP サービスが与える影響について調査を行う。調査するにあたり、RESIP ホストへの影響は RESIP 事業者によって秘匿されており調査が難しいため、RESIP 利用者に注目して、サービスを用いることによる影響を明らかにすることを目的とする。調査対象の RESIP プロバイダは規模の大きさや先行研究 [1][2] で調査状況を踏まえ、Bright Data^{*1}と ProxyRack^{*2}とした。本稿では、RESIP を利用時に RESIP ホスト、プロバイダによって位置情報、ターゲット広告与える影響に差があるかの調査を報告する。

2 RESIP サービスとターゲット広告

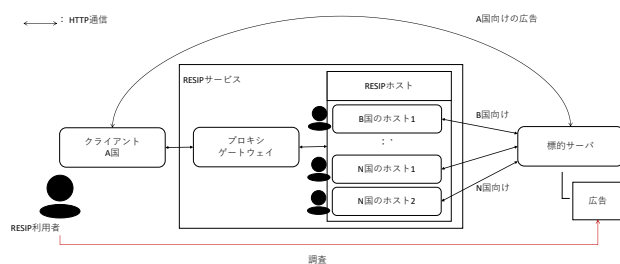


図1 RESIP サービスのシステム構成図

図1は RESIP サービスを用いた広告調査の構成図である。RESIP サービスは住宅用 IP によって通信を中継するサービスであり、クライアント、プロキシゲートウェイ、住宅用ホストから構成されている。RESIP 利用者がプロキシを利用する場合、ゲートウェイは RESIP 利用者から来た通信を設定に応じて異なる RESIP ホス

(a) プロキシなし (b) プロキシあり

図2 ターゲット広告の例

トへと割り当てる。標的サーバからの応答は割り当てられたホストを経由して RESIP 利用者へ中断される。

ターゲット広告とは、ユーザの web ページ閲覧履歴などを基に配信されるユーザを標的にした広告である。図2(a)のターゲット広告はプロキシを経由すると図2(b)になった。広告の内容だけでなく、広告に使用されている言語も異なっている。ターゲット広告のうち、RESIP ホストによって内容が変化するものを標的と呼び、全ターゲット広告数に対する割合を標的率と定義する。

3 実験

3.1 目的

本実験は RESIP サービスを用いることで位置情報とターゲット広告に影響を与えるのかを調査し、RESIP サービスの影響を判断させることが目的である。

3.2 方法

3.2.1 実験1：位置情報の調査

現在地を取得したとき実際の現在地から変化するのか Google Maps(<https://www.google.co.jp/maps>) を使用し調査する。調査は各プロキシプロバイダにおいて40個の RESIP ホストを利用し、合計80個の RESIP ホストに対して2021年9月9日から11日にかけて行った。Bright Data で利用した RESIP ホストはアメリカ合衆国、カナダ、ウクライナの3か国に位置する RESIP ホストを各10個、日本、韓国の2か国に位置する RESIP ホストを各5個である。ProxyRack は地域を指定せず40個の RESIP ホストを利用した。

3.2.2 実験2：ターゲット広告の調査

ターゲット広告へプロキシの影響があるか、5つの web サイトを対象として2つのプロキシプロバイダを用いて調査する。対象とした5つの web サイトを表1に示す。Bright Data の RESIP ホストは実験1と同じ条件で行った使用し、ProxyRack は実験1と全く同一の RESIP ホストを用いて調査を行う。

プロキシの影響を受けている標的のターゲット広告は、次の3つの条件のいずれかに当てはまるものと定義する。明確に満たしていない、判断が難しいものについてはプロキシの影響がないものとする。

Research on location-based and targeted advertising using the Residential IP Proxy service

†Hikari Fukuda, Tatsuya Ikubo, Hiroaki Kikuchi, School of Interdisciplinary Mathematical Science, Meiji University.

*1 <https://brightdata.com/>

*2 <https://www.proxyrack.com/>

表1 調査対象の web サイト

| web サイト | 本実験での名称 | URL |
|--------------|---------|---------------------------|
| Yahoo! JAPAN | yahoo | https://www.yahoo.co.jp/ |
| 神ゲー攻略 | game | https://kamigame.jp/ |
| folk | folk | https://folk-media.com |
| ロケットニュース 24 | rocket | https://rocketnews24.com/ |
| Smartlog | smart | https://smartlog.jp/ |

表2 位置情報の変化数

| | Bright Data | ProxyRack |
|--------|-------------|-----------|
| 現在地の変化 | 0/40 | 38/40 |

1. 日本語以外の言語で書かれている
2. 日本語表記であり、海外在住者向けの広告である（日本商品の通信販売や日本在住と同様のインターネット環境の構築など）
3. 日本語表記であり、現在地と異なる地名が入っている

以下に調査の手順について示す。調査は 2021 年 8 月 22 日から 9 月 11 日にかけて手動で行った。

1. RESIP を経由し表 1 のサイトにアクセスして、ターゲット広告の個数、RESIP 利用により表示内容が変化したターゲット広告の数を記録する。
2. 1 を各 RESIP アドレスに対して 5 つの web サイトで行い、繰り返す

3.2.3 実験 3：自動プログラムによる観測

実験 2 をプログラムにより自動化した調査を 2021 年 11 月 12 日から 16 日までの 4 日間で行った。

本プログラムは node.js のライブラリである puppeteer[3] によって RESIP サービスの設定、ブラウザ操作、画像取得を自動化する。実験 2 との差異は、手順 1 で遷移した web サイトの全体のキャプチャを行うことである。これを自動で行うことにより、web サイトが読み込まれるまでの待機時間を大幅に削減することができた。

3.3 結果

3.3.1 実験 1：位置情報の調査結果

位置情報についての実験 1 の結果を表 2 に示す。位置情報が現在地と異なっていたサイト数を示す。

表 2 より、Bright Data では全ての RESIP ホストにおいて現在地の取得がでなかった。ProxyRack では 403 Forbidden のエラーメッセージが返されるサイトが 4 個あるが、それ以外の RESIP ホストにおいて現在地が異なった。Bright Data では、「現在地を特定できませんでした」という Google Maps のエラーが返される。

3.3.2 実験 2 と実験 3 のターゲット広告の調査結果

表 3 は標的広告数の割合について示している。

Bright Data と ProxyRack 共に手動で行った実験 2 に比べて実験 3 の標的率が低く、特に Bright Data で

表3 調査結果：標的率

| | Bright Data | | ProxyRack | | 平均 | |
|--------|-------------|------|-------------|------|-------------|------|
| | 手動 | 自動 | 手動 | 自動 | 手動 | 自動 |
| yahoo | 0.05 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 |
| game | 0.74 | 0.02 | 0.48 | 0.21 | 0.65 | 0.10 |
| folk | 0.82 | 0.02 | 0.68 | 0.43 | 0.74 | 0.19 |
| rocket | 0.54 | 0.05 | 0.70 | 0.12 | 0.61 | 0.09 |
| smart | 0.83 | 0.02 | 0.52 | 0.40 | 0.68 | 0.20 |
| 平均 | 0.68 | 0.02 | 0.55 | 0.27 | 0.62 | 0.13 |

は平均標的率が 97% 下がった。rocket を除き標的率は、手動では Bright Data が高く、自動では ProxyRack が高かった。

4 評価

4.1 Residential IP Proxy プロバイダの比較

位置情報の変化に関しては、表 2 より Bright Data では現在地が表示されなかったことから、プロバイダにより位置情報が変化することがわかった。

ターゲット広告の調査について、表 3 より Bright Data を対象にした調査が最も標的率が高くなった。

4.2 考察

ターゲット広告の調査において yahoo の変化の割合が著しく低い理由は、yahoo が日本の IP アドレスからの通信にのみターゲット広告を表示することが原因であると考えられる。

5 おわりに

表 2 より、ProxyRack では 40 個のうち 38 個の RESIP ホストで位置情報が変化し、Bright Data においては位置情報サービスが機能していなかった。ターゲット広告に関しては表 3 より、BrightData と ProxyRack のいずれの調査においても標的が確認された。本調査では、上記 2 点において位置情報とターゲット広告の変化が確認されたことを明らかにした。

本稿では RESIP サービスが RESIP 利用者へ与える影響を調査した。今後は RESIP 利用者が RESIP ホストへ与える影響を調査することを課題とする。

参考文献

- [1] Xianghang Mi, et al., Your Phone is My Proxy: Detecting and Understanding Mobile Proxy Networks, NDSS Symposium 2021, pp.1-18, 2021.
- [2] 半澤, “Residential IP Proxy サービスに悪用される住宅用ホストの調査”, 2020 年度菊池研究室修士論文, 2021.
- [3] Alex Rudenko, “GitHub - puppeteer/puppeteer at v13.0.0” (2021 年 12 月参照)