

被害額等の算定に利用する データについて

(一社)日本インターネットプロバイダー協会

副会長兼専務理事 立石 聡明

2018年9月19日

被害額の算定について

- 2018年4月、政府の犯罪対策閣僚会議で発表された「インターネット上の海賊版サイト に対する緊急対策」において
 - 【「財産権であることを もってすなわち回復可能」と断じるのではなく、こうした特に悪質な 海賊版サイトに係る状況を勘案した上で、事例に即した具体的な検討 が求められる。】とあり
 - 「保護されるべき著作物が公開されることによりどの程度回復困難な損害を生じ得るか」という項目が掲載されているが、本タスクフォースにおいて検討されたとは思えない。
- またこれまで被害額の算定根拠とされていた、SimilarWeb社のデータに関して大きな疑問を感じたために調べたところ、Ad Fraud等との関連について次頁以降の様なものが得られた。

SimilarWeb社が買収したエクステンション ”Stylish“が違法に個人データを採取していた

- <https://robertheaton.com/2018/07/02/stylish-browser-extension-steals-your-internet-history/>
- 上記のレポートによると、SimilarWeb社がChromeのエクステンションである”Stylish”を買収したと2017年1月に発表(Chromeに限らない)
- このエクステンションは200万人以上が利用
- このエクステンションは、利用者の全てのブラウザ閲覧履歴を無断でSimilarWeb社に送信。
- 閲覧履歴は匿名化しているとしているが実際は各個人に紐付けられた形でSimilarWebに送信されていた
 - Base64で2回エンコードするなどして発見されないよう送信データを偽装していた
- また閲覧履歴だけでなく、e-mailや認証トークンまで盗み出していたことが分かっている

SimilarWeb社について

- 2016年6月7日ネットギーク等の記事でSimilarWebのデータに信憑性がないことなどで炎上していることを報道
 - 「【炎上】 SimilarWebの信憑性のないデータに各メディア責任者が網抗議」
<http://netgeek.biz/archives/74929>
- 2017年1月SimilarWebがStylishを買収
 - Stylishのユーザは200万人を超えており、インパクトは非常に大きい
- このことから SimilarWebは2016年6月の炎上を真摯には受け止めておらず、さらなるマルウェアを配布する為にツール・ベンダーを買収して個人データの抜き取りをはじめとする違法行為等を繰り返しているようである

Stylishとその機能について

- Chrome等のエクステンションで、本来は任意のページを自分好みのスタイルに変更するためのものでユーザは200万人以上
- 利用ユーザ全てのブラウザ履歴をSimilarWebや関連会社に送信
 - Base64で2回エンコードするなどして、文字列を粉飾し簡単には分からないようにして送信 ← 悪意あることを証明
- また、これらがBotとなって機械的なアクセスを生んでいる可能性を指摘する記事も
- ChromeやFireFoxのエクステンションやアドオンとしてはSpywareあるいはMalwareと認定されてそれぞれの公式サイトから削除されている

SimilarWeb社のデータ利用に関して

- SimilarWebに関しては

1. *FTCは、SimilarWebを事例研究として実名入りで注意喚起*

https://www.ftc.gov/system/files/documents/public_comments/2017/11/00043-141901.pdf

個人情報 を無断で自社に送信していると断言

2. *ハーバードビジネススクールEdelman教授が問題視*

SimilarWeb社を利用すること自体に重大な責任があると指摘

- このような会社のデータを

1. 政府の資料としていることに大きな疑問を感じると共に、
2. 今後大きな問題を招く可能性を孕んでいないのかと感じる

Ad Fraudの仕組み

- 「漫画村」で問題となり始めたAd Fraudには、もっと複雑な仕組みが存在する
- Malwareに感染したりBotと化したブラウザが、ユーザの意思とは関係なく、様々なサイトを渡り歩き、PVを稼ぐなどソフトウェアがアクセス数を稼ぐ
- 広告のトラフィックを得る業者も存在
- いわゆる広告詐欺の手法としては下記の様なものとされている
 - **隠し広告 (Hidden Ads)**、**詰め込み広告 (Ad Density)**、**オートリロード (Auto Refresh)**、**広告経路による第三者からのトラフィック獲得 (Sourced Traffic)**、**トラフィックエクスチェンジ経路による第三者からのトラフィック獲得 (Sourced Traffic)**、**オークションのURL偽装 (Falsely Represented)**、**データセンターからのトラフィック流入 (Data Center / Proxy)**、**クッキーの不正操作 (Cookie Stuffing)**、**不正な広告挿入 (Ad Injection)**、**ウイルス・マルウェア・アドウェアなどによる端末の乗っ取り (Malware, Adware, Hijacked Device)**、**ブラウザの不正操作による広告閲覧を生成 (Imp/click Bot, Retargeting Fraud)**、**不正クリック発生によるアトリビューションの汚染 (Click Stuffing)**、**不正クリックによるコンバージョンの乗っ取り (Click Injection)**、**バックグラウンドでの広告閲覧 (Background)**

FTCやその他の団体、教授が独自のレポート

- https://www.ftc.gov/system/files/documents/public_comments/2017/11/00043-141901.pdf
 - **ブラウザのエクステンションが閲覧履歴を違法に漏洩している件について検知**
 - 42エクステンションのうち、オプトインは1つのみで多くはまともな規約もない
 - “A CASE STUDY OF A LARGE HISTORY DATA COLLECTOR”の例としてSimilarWebを事例紹介
- <https://www.adweek.com/digital/six-companies-fueling-online-ad-crisis-150160/>
 - 0.5Pennyからアクセス数を売っている広告会社(Botから生成したアクセスの可能性が高い)
 - AdOn, Adknowledge, eZanga, Jema Media, MGID and BlueLink Marketing
 - これらの広告トラフィック販売会社のデータを元に分析を行っている3大会社が
 - SimilarWeb, eZanga, BlueLink
 - **Harvard Business School professor Ben Edelmanによれば、Smartmomstyleなどこの会社を利用すること自体に重大な責任があると指摘**
- <https://www.screamingfrog.co.uk/how-accurate-are-website-traffic-estimators/>
 - SimilarWebは、データを4割増しに
- <https://www.bloomberg.com/features/2015-click-fraud/>
 - Bloombergの上記サイトにはAd Fraudに関する詳細なレポートがあり、やはりSimilarWebについても言及

その他、Ad FraudがBot等でアクセス数を増やしたり、個人情報抜き取っていることを報告する記事は無数に存在する

- BloombergがAd Fraudについて詳しく報告
 - <https://www.bloomberg.com/features/2015-click-fraud/>
- Firefoxの人気アドオン「Stylish」はユーザーのインターネット履歴をこっそり収集していた。
 - <https://gigazine.net/news/20180705-stylish-steals-your-internet-history/>
- 人気のFirefox拡張機能「Stylish」がポリシー違反でブロック、すべての閲覧履歴を収集か
 - <https://forest.watch.impress.co.jp/docs/news/1131143.html>
- SimilarWebの訪問者データは、いったいどこから取得しているのか？
 - <https://nelog.jp/similarweb-data-is-from>
- Firefoxの人気アドオン「Stylish」はユーザーのインターネット履歴をこっそり収集していた
 - https://article.auone.jp/detail/1/3/7/48_7_r_20180705_1530755712280156
- 365,000の広告インプレッションのうち、なんと57%が「自動化されたコンピュータプログラムによる閲覧」と報告
 - <http://global-adtech.jp/blog/1066>

漫画村のサイトにアクセスしていた人やそのPVは一体どれほどなのか？

- 官邸発表の資料や知財本部の資料ではSimilarWebのデータが、被害額の算定基準になっているが、果たしてこの数字は信用するに値するか？

「2018年2月の訪問者数はのべ1億6千万人、日本からのアクセスだった。」

知財本部採用の訪問者数…一億六〇〇〇万



下記サイト等の分析では、SimilarWebを含む複数のアクセス数分析会社は、その多くが水増ししていると指摘。下記のサイトでは、Google Analyticsの実データと比較して4割程度水増ししている可能性があるとの分析。
となると、この場合4000万以上は水増しされていた事になる。
<https://www.screamingfrog.co.uk/how-accurate-are-website-traffic-estimators/>

別のレポートではBot(非人間)によるアクセスが場合によっては7割に上るという分析結果。
ここでは、仮に半分としても、人間による実アクセス数は6000万弱が人間以外のアクセス。
(別のレポートではBotが1時間に200万回のアクセスするという報告もある。)

「7割」を控えめに「5割」として計算すると人によるアクセス数は6000万弱になる。
仮に、この数値(6000万)が実数に近かったとしても、さらなる疑問が生じる → 次頁

実際は6000万人より(あるいはそれよりずっと)少なかった可能性も！？

「漫画村」の被害額の想定について

- 被害額の想定には様々あり、裁判例でもいろいろあるようだが、仮に知財本部の出した方法で計算するとしても元データが大幅に異なると
 - 期間中の合計訪問者数
 - 訪問者数6億1989万人のうち4割がSimilarWeb社に水増しされているとすると
 $6億1989万人 \div 14 \times 10 = 約4億4278万人$
 - そのうち仮に少なめに見積もって、半数がBot(非人間)だとすると、
 $約4億4278万人 \div 2 = 約2億2139万人$
 - 想定被害額が2億2139万人 × 515円 = 約1140億円となる
 - また同試算によると17.42分で1冊読むものと仮定しているが他のデータでは30分から1時間という数値もある
 - 他のデータの平均より短めに約35分で読むと仮定すると単純計算で**被害額は570億円程度**となる。

算定根拠になっているデータの取り扱いについて

- 更に「**著作権法上の損害の推定規定**」によれば
 - 権利者の単位数量あたりの利益額 × 譲渡数量又は公衆送信で作成された複製物（受信複製物）の数量
となっているため、実際の裁判で損害の算定をする際には、更に小さな数字となる可能性は非常に大きいと言える
 - 試算として1冊あたりの利益を3割程度であるとした場合
 - **出版社の逸失利益：570億円 × 30% = 171億円**
 - 更に裁判となった際に、元の数字である「約2億2139万人」の全てが認められる可能性は低いため、更に小さな数字となる可能性も出てくる。
 - 「有料」の場合には、「無料」の場合の閲覧数とは経験則上同じではない
 - 今回の漫画村では、多くが小中学生であるとの指摘もあり、彼らが有料だった場合にどれくらい買うのかを推定する必要もあるのではないか
- FTC等のレポートからも、SimilarWeb社のデータを掲載することは極めて不適切であると言わざるを得ない。
- よって推定されている約3000億の被害想定額はとりまとめ案から削除するのが適当ではないか