

セッションリプレイサービスからの個人識別性と国内外サイトにおけるプライバシーポリシーでの公表状況

梶間大地^{1,a)} 菊池浩明^{2,b)}

概要: セッションリプレイサービスは、ユーザがウェブサイト上で行なった操作を記録するサービスである。これはユーザ体験の向上やエラー原因の特定に利用されることが想定されている。しかし全ての導入サイトが、ユーザに対してセッションリプレイサービスを利用してユーザのマウス履歴などの情報を取得していることを公表しておらず、ユーザがプライバシーの侵害を感じる懸念される。

そこで、本研究ではユーザがマウスやキーボードの操作を取得されることに対してどのように感じているのかをクラウドソーシングを用いたアンケート調査から明らかにする。次に、これらの操作履歴から、特定のユーザを識別できるかを実験的に検証する。最後に、国内外サイトにおいて、セッションリプレイサービスの導入状況とプライバシーポリシーの適切な公表に関する調査を行い、その結果を報告する。

Personally identifiable from session replay service and domestical and abroad disclosure status in privacy policy

Abstract:

Session replay is a service for recording every active event that a user performs on a website. It is supposed to be used for improving the website user experience and identifying failure on the website. However, not all of the sites having this service disclose to their users that they use the session replay service to obtain information such as the user's mouse history, and hence there is a concern that users may feel their privacy being violated.

In this paper, we investigated how users feel about their mouse and keyboard operations being captured through a questionnaire survey. In addition, we conducted a survey about the deployed session replay services and whether they are appropriately disclosed in privacy policies in both domestic and abroad websites.

1. はじめに

セッションリプレイサービスは、ウェブサイト上でユーザが行ったマウスやキーボード操作を記録するサービスである。このサービスは、ユーザの記録からウェブサイトのどのような箇所で操作に迷っているのかや、どのような操作をしたときにエラーが発生するのかを記録から特定することで、ユーザ体験の向上やエラー原因の特定に利用することを目的としている。

代表的なセッションリプレイサービスとして、Clarity[9], Hotjar[10], Yandex[16] が挙げられる。これらのサービス

は著名なウェブサイトでも利用されており、日本では新聞社のウェブサイト^{*1*}やゲーム攻略サイト^{*3}などが導入を行なっている。

しかし、これらのサービスを導入しているサイトの全てがユーザに対して情報を取得していることを公表しているわけではない。例えば、我々が調査した国内 300 サイトにおいては、少なくとも 56 件のサイトがセッションリプレイサービスを用いてマウス履歴などを取得していることとその目的をプライバシーポリシーで言及していなかった。従って、多くの利用者は自分のマウス履歴などを取得されていることに気がついていないことが懸念される。

事業者が公表を行わないことによって、ユーザにとって機微な情報がユーザが知覚されることなく送信されてしま

¹ 明治大学大学院先端数理科学研究科
Nakano, Nakano-ku, Tokyo 164-8525, Japan

² 明治大学総合数理学部
Nakano, Nakano-ku, Tokyo 164-8525, Japan

a) cs222014@meiji.ac.jp

b) kkn@meiji.ac.jp

*1 <https://www.yomiuri.co.jp/>

*2 <https://mainichi.jp/>

*3 <https://gamewith.jp/>

うことがある。加えて、マウスの操作やキーボードの操作履歴は個人を識別することが可能である情報であることが泉らによって示されている [4][5][6][7]。したがって、セッションリプレイサービスで取得される情報は個人に関する情報*4に当たる可能性がある。また、違法性はなくとも自分のマウスの動きやキーストロークが知らないうちに取得されていることに対して、否定的な感情を持つユーザも少なくないと考えられる。しかも、これらのセッションリプレイサービスによって自分のマウス履歴などが取得されているかどうかを判断するためには、ウェブページのソースコードを解析する必要があり、専門性知識を持たない利用者にはセッション情報取得を検出することが非常に困難である。従って、本来ならば、サイト事業者は個人に関する情報の取得の事実と利用目的について、適切な公表を行いユーザに対して情報の取得の同意を取ることが望ましいと考える。

そこで本研究では、以上の利用者の意識に対する仮説を検証するために、まず、(1) クラウドソーシングによる 210 名の利用者を対象とした、セッションリプレイサービスの認知度と、セッションリプレイサービスに取得される情報に対する許容度のユーザ調査を行う。次に、セッション情報からの利用者識別リスクを定量化するために、マウス履歴を取得する独自のセッションリプレイサービスを研究室内に試験実装を行い、(2) マウスの移動履歴からどれだけ正確に利用者が識別できるか、試験的にセッションを取得するサイトを実装して個人識別性を明らかにする技術的な実験評価を実施する。最後に、(3) 国内外の主要な約 17,000 ウェブサイトにおけるセッションリプレイサービスの導入状況を調査し、セッション情報取得の利用目的などの情報を利用者に適切に公表している割合を明らかにする。その結果を報告する。

2. 関連研究

2.1 セッションリプレイサービスによるプライバシー保護の研究

2020 年、Gunes らは 50,000 のウェブサイトに埋め込まれたサードパーティのスクリプトによる情報の流出について OpenWPM を利用した調査を行った [2]。FullStory, UserReplay, SessionCam, Hotjar, Yandex, Smartlook の 6 つのセッションリプレイサービスの DOM の収集によって起こる情報の流出を明らかにした。[2] によると、DOM の収集によってクレジットカード情報やパスワードなどの認証情報や機密情報をセッションリプレイサービスのサーバに送信しているウェブサイトが存在する。

Xiufen らは、アジア、ヨーロッパ、北米、中南米、アフ

リカ、オセアニアの 152 カ国・地域の 19,483 の病院のウェブサイトにおけるセキュリティとプライバシーについて調査を行った [3]。その結果、Hotjar が 469 サイト、Yandex が 226 サイト、fullstory が 4 サイトで導入されていることを明らかにした。彼らは、プライバシーポリシーにおいて個人情報を第三者に開示しないことを明示しているにもかかわらず、電話番号やメールアドレスなどをセッションリプレイサービスにサーバに送信しているサイトが存在することを明らかにした。

[2][3] では、セッションリプレイサービスによる情報流出のみ調査を行っており、セッションリプレイサービスの導入状況、ユーザに対する公表状況は明らかになっていない。

2.2 操作による個人識別研究

マウスとキーボード操作から個人を識別することが可能であることがよく知られている。

泉らは、コンピュータ操作時のマウスの移動方向のパターンから個人を識別する方式を提案した [4]。5 人のユーザから取得したデータに対して識別実験を行った結果、再現率、適合率ともに 70% 以上の精度で識別が可能であることを明らかにした。

須田らは、コンピュータのスリープ復帰時におけるマウス操作から個人を識別する手法を提案した [5]。マウス操作時の軌跡から移動距離や軌跡の長さ、所要時間、カーソル速度の平均、中央値、標準偏差、カーソル加速度の平均、中央値、標準偏差などの特徴量に基づいて個人を識別する。実験結果より、被験者が 4 人の場合では平均等価エラー率 2.5%、12 人の場合では、10.0% という精度で個人を識別できることを明らかにした。

木村らは、収集したキーボードとマウス操作のログデータに深層学習を用いることで、操作者を特定する手法を提案した [6]。実験結果より、任意の 2 つの操作ログが同一の操作者によるものであるかを 90% 以上の精度で判別することができることを明らかにした。

渡邊らは、スマートフォンにおけるタッチ操作の特徴量からの個人識別について研究を行った。その結果、指の使用領域や移動距離、速度などの操作特徴を組み合わせることによって個人が識別可能であることを示した。

しかし、[4][5][6][7] では、特殊な操作時のログからの個人識別について研究を行っており、通常時のマウス操作からの個人識別については不明である。

3. セッションリプレイサービスの認知度調査

3.1 実験方法

2024 年 1 月に、クラウドソーシングサービスであるクラ

*4 個人情報保護法で規定されている個人情報ではないが、いわゆる、パーソナルデータとして、個人の識別性を有する潜在性がある情報。

ウドワークス*⁵を用いて参加者 210 人（男性 98 人，女性 112 人）に対してアンケート調査を行なった。参加者の属性を表 1 に示す。

アンケートの内容は下記の 4 つのパートからなる。実際に用いたアンケート設問を付録 A.1 に示す。^{*6}

- (1) 参加者個人属性（質問 1-6）
- (2) 架空サービスに対する許容度（質問 7-11）
- (3) 参加者のセキュリティ意識（質問 12, 13, 15）
- (4) ウェブサイトからのセッションリプレイサービスの有無の知覚可能性（質問 16-19）

(1) では，参加者の属性について質問をした。(2) では，ユーザがどのような要素を取得されることに対して気持ち悪さを感じるか明らかにするために，表 2 に示す 6 種類のサービスを設定し，そのサービスに対してどのように感じるか，また，そのサービスを利用するかを質問した。(3) では，被験者のセキュリティ意識の強さと IT 技術に対する知識について質問を行なった。(4) では，セッションリプレイサービスの導入されているサイトとされていないサイト 2 つの計 3 つのサイトを与え，どのサイトで導入されているか知覚できるかどうかを調査する質問を行なった。加えて，セッションリプレイサービスを知っているか，マウスやキーボード操作を取得されることに対してどのように感じるかについて質問を行なった。

ユーザのサービスに対する許容度が IT スキル，セキュリティ意識に依存するかどうかを確かめるために，独立性のカイ二乗検定を行なった。IT スキルは，4 未満と 4 以上の 2 群に分割し検定を行なった。セキュリティ意識については，各質問ごとに点数づけを行った。5 点を基準に 5 点未満と 5 点以上の 2 群に分割し検定を行なった。

3.2 結果

架空のサービスに対するユーザの許容度について，アンケートの結果を表 2 に示す。

最も許容度が高いのは，11 のページデザインを目的としたサイトであり，81%(115+56/210) の利用者が許容していたのに対し，目的を示さずに同様のスクロール操作を取得しているサービス 6 の許容度は，48%(34+68/210) に下がる。

質問 7 と 11 はどちらもユーザのマウス情報を取得するサービスに関する質問である。しかし，質問 7 の商品推薦の許容度は，61%(64+65/210) に下がる。

IT スキルと許容度，および，セキュリティ知識と許容度についての独立性の検定結果を表 3 表 4 示す。IT スキルは，全ての質問において有意な差は見られなかった。セキュリティ意識では，目的の記載のないサービスについて

表 1 参加者の属性

	属性	人数	%
年代	20 代以下	18	9
	30 代	84	40
	40 代	71	34
	50 代	30	14
	60 代以上	7	3
性別	男性	98	47
	女性	112	53
	その他	0	0
ウェブサイト閲覧頻度	毎日	199	95
	週 5~6 日	10	5
	週 3~4 日	1	1
	週 2 日以下	0	0
IT スキル	5	2	1
	4	35	17
	3	104	50
	2	44	21
	1	25	12

有意な差が見られた。設問 6, 8, 10 の架空サービスに対して，セキュリティ意識が高い人の許容度が高く，0.05 の有意水準を超えている。セッションリプレイサービスを初めとするクライアント識別が広く行われていることを既に知っている為と考えられる。また，セッションリプレイの設置の検出については，77%が分からないと回答しており，検出が技術的に困難であるとする我々の仮説通りであった。

以上により，セッションリプレイサービスに対する利用者の否定的な感情は十分に大きいにも拘らず，その検出は困難であることが明らかにされた。

セッションリプレイサービスの認知について，結果を表 5 に示す。74.8% (157 人/210 人) がサービスの概要も名称も知らないと回答した。マウスやキーボード操作を取得されることに対しては，62.3% (131 人/210 人) が取得されることに対して否定的であると回答した。

3.3 考察

質問 7 と 11 の結果より，ユーザは取得される情報によって許容度が変化するのではなく，情報の使用目的によって許容度が変化することが明らかになった。これは情報の利用目的がユーザ自身に関係ある場合，ユーザは気持ち悪いと感じ，許容度の低下に繋がると考えられる。

4. ブラウザ上における個人識別性

4.1 概要

本大学の学生 15 人のウェブサイト利用時のマウスとスクロールの操作ログを取得した。

取得したログに対してランダムフォレストを用いて，2 つの操作ログが同一の操作者であるか判定を行なった。識別を行なった際の特徴量は先行研究 [5] を参考に表 6 の特

*⁵ <https://crowdworks.jp/>

*⁶ 設問 14 は，被験者が正しく設問を読んで回答しているかを確認するための自明な問である。

表 2 架空のサービスに対するユーザの許容度の調査結果

番号	質問	気にしない	気持ち悪いが利用する	気持ち悪く利用しない
6	あなたのスクロール操作が全て記録されるウェブサービス	34	68	108
7	あなたのマウスの動きから悩んでいる商品を推測し類似したおすすめの商品を推薦するショッピングサイト	64	65	81
8	あなたの閲覧時間が記録されているウェブサイト	79	65	66
9	あなたのキーボード入力内容から最適な広告を配信するウェブサイト	67	67	76
10	あなたがそのページを何回閲覧したかを記録しているウェブサイト	80	70	60
11	あなたのマウスの動きからページのデザインを改善しているウェブサイト	115	56	39

表 3 ユーザの許容度と IT スキルのカイ二乗検定結果

質問	IT スキル		統計量	p 値
	あり	なし		
6	4	30	0.9579	0.3277
7	10	54	0.2522	0.6155
8	14	65	0.0009	0.9759
9	11	56	0.0978	0.7545
10	14	66	0.0013	0.9717
11	22	93	0.4001	0.5271

表 4 ユーザの許容度とセキュリティ意識のカイ二乗検定結果

質問	セキュリティ意識		統計量	p 値	< 0.05
	あり	なし			
6	27	7	7.5721	0.0059	***
7	43	21	3.1260	0.0771	
8	56	23	8.5108	0.0035	***
9	41	26	0.3881	0.5333	
10	57	23	9.1860	0.0024	***
11	70	45	0.8037	0.3700	



図 1 実験 1 に使用したウェブサイト

微量を用いて、100 回識別を行い精度、適合率、再現率、F 値の平均値を算出した。

4.2 実験 1

図 1 に実験に使用したウェブサイトを示す。枠内をクリックしながらマウスを移動させることで、イラストを描出できるウェブサイトを構築した。マウスを使って星のイラストの描出を依頼し、その際のマウスの操作ログを取得した。

4.2.1 実験 2

ウェブサイト*7から、特定年度の研究室の所属人数のような質問の答えを探し出すタスクを依頼した。その際のマウスとスクロール操作を取得した。

4.2.2 結果

結果を表 7 に示す。適合率、再現率ともに 0.27 以上で識別可能である。与えられたマウス履歴の特徴量について、15 人の中からランダムに識別が正しい確率は 1/15 である。従って、0.27 はそれより 4 倍高い精度である。

4.3 考察

先行研究 [5] では 12 人の被験者から取得した操作ログに対して識別を行なった結果、90.3%の精度で識別が可能であったのに対し、本実験では精度が低い結果となった。取得した操作ログに対する前処理が不十分であったことが精度の低下に繋がったと考えられる。本実験では、取得した操作ログに対して前処理を行っていない。効果的な前処理を行うことで、精度が向上する可能性があると考えられる。

以上より、マウスの移動履歴に基づいて利用者を識別して追跡するリスクは、ランダム推定よりも十分に高い精度で可能であることと、丁寧な前処理を施すことで更に識別精度が向上することが明らかになった。

5. セッションリプレイサービスの導入状況調査

5.1 調査方法

下記に調査の手順を示す。調査は Python を用いて自動で行なった。

- (1) ブラウザ自動化フレームワークの Selenium[23] を用いて、対象サイトにアクセスする。
- (2) アクセスした際にブラウザから外部サーバに送信される HTTP リクエストをデータベースに格納する。
- (3) 収集したログから既知のセッションリプレイサービスの URL から JavaScript のスクリプトを読み込んでいるサイトを特定する。
- (4) Norton Safe Web[25] のカテゴリに従いサイトの業種

*7 <https://www.kikn.fms.meiji.ac.jp/kajima/2023/exp/html>

表 5 セッションリプレイサービスの認知調査結果

質問	回答	人数	%
あなたはセッションリプレイサービスを 知っていますか	知っている	2	1.0
	名称は知っているがサービスの概要は知らない	11	5.2
	名称は知らないがそのようなサービスがあることは知っている	40	19.0
	知らない	157	74.8
マウス操作やキーボード操作が 取得されることに対して どのように感じるかお聞かせください	5. 取られたくない	49	23.3
	4. あまり取られたくない	82	39.0
	3. どちらでもない	46	21.9
	2. あまり気にしない	26	12.4
	1. 気にしない	7	3.3
下記の 3 つのウェブページのうち、1 つのサイトでは セッションリプレイサービスを利用しています。 このうち、どのサイトで利用しているか教えてください	正答	17	8.1
	誤答	31	14.8
	わからない	162	77.1

表 6 識別に使用した説明変数

説明変数	実験 1	実験 2
マウス操作に要した時間	✓	-
カーソル速度の平均	✓	✓
カーソル速度の中央値	✓	✓
カーソル速度の標準偏差	✓	✓
軌跡の x 軸の最大値	✓	✓
軌跡の x 軸の最小値	✓	✓
軌跡の y 軸の最大値	✓	✓
軌跡の y 軸の最小値	✓	✓
描出した軌跡の x 軸の最大値	✓	-
描出した軌跡の y 軸の最大値	✓	-
スクロール速度の平均	-	✓
スクロール速度の中央値	-	✓
スクロール速度の標準偏差	-	✓
総移動距離	✓	-
データ取得数	✓	-

表 7 操作ログによる識別結果

	実験 1	実験 2
Accuracy	0.3697	0.3618
Precision	0.4333	0.2786
Recall	0.3502	0.3092
F_1	0.3407	0.2681

を分類する。

図 2 に Selenium から取得できる通信ログの例を示す。これは、Selenium がネットワークに関するイベントが発生した際に出力されるものである。method には発生したイベントの種類が出力される。“Network.requestWillBeSent”とは、Selenium でリクエストを送信するときに発生するイベントである。このイベントが発生した際に、リクエストの送信先が既知のセッションリプレイサービスの JavaScript の URL と一致した場合に、サービスを利用していると仮定した。

Norton Safe Web とは、NortonLifeLock 社の提供するウェブサイトの安全性やカテゴリに分類するサービスである。本調査では、カテゴリ分類を自動で行い、国内外サイ

トのカテゴリの違いを明らかにするために利用した。

```

{"level": "INFO",
 "message":
  "{
    \"message\": {
      \"url\": \"https://gamewith.jp/\",
      \"params\": {
        \"documentURL\": \"https://gamewith.jp/\",
        \"referrer\": \"https://gamewith.jp/\",
        \"referrerPolicy\": \"strict-origin-when-cross-origin\",
        \"requestId\": \"66010.2284\",
        \"timestamp\": 167344572.8407463,
        \"type\": \"Script\",
        \"wallTime\": 167344572.8407463,
        \"webView\": \"033A08C0D916057C62A1B436638B30V\"
      }
    }
  }
}

```

図 2 Selenium から収集した通信ログの例

5.2 調査対象

調査対象とするセッションリプレイサービスは、Microsoft Clarity[9], Hotjar[10], mouseflow[11], crazyegg[12], ContentSquare[13], luckyorange[14], fullstory[15], Yandex[16], Dynatrace[17], Glassbox[18], Smartlook[19], Foresee[20], Inspectlet[21], LogRocket[22] の 14 のセッションリプレイサービスとした。

ウェブサイトは、トップウェブサイトランキングを提供している Tranco[8] の 2023 年のデータセットのうち、事前に GET リクエストを送信しステータスコード 200 (正常) が返ってきた 17,582 サイトを対象とした。

5.3 結果

2023 年 11 月における国外サイトにおけるセッションリプレイサービスの導入状況と調査結果を表 8 に示す。国内では、Clarity が 74% のシェアであるのに対し、国外では 22.9% で Hotjar, Yandex に続く第 3 位である。一方、国外での 1 位の Hotjar は国内では 9.4% に過ぎず、Clarity に比較して著しく利用が少ない。

Norton Safe Web を利用したカテゴリ毎の調査結果を

表 8 セッションリプレイサービスの導入状況

サービス名	国内 [1]		国外	
	件数	%	件数	%
Clarity	702	74.4	818	22.9
Hotjar	89	9.4	1127	31.5
Mouseflow	68	7.2	86	2.4
Yandex	4	0.4	791	22.1
contentsquare	20	2.1	88	2.5
Crazyegg	40	4.2	380	10.6
Dynatrace	2	0.2	39	1.1
foresee	1	0.1	35	1.0
fullstory	6	0.6	110	3.1
glassbox	2	0.2	10	0.3
inspectlet	0	0	31	0.9
logrocket	0	0	6	0.2
luckyorange	7	0.7	33	0.9
Smartlook	2	0.2	21	0.6
合計	943		3575	

表 9 Norton Safe Web による国内外の導入サイトの上位カテゴリ

カテゴリ	国内		国外	
	件数	%	件数	%
テクノロジー/インターネット	150	15.9	857	24.0
ビジネス/経済	181	19.2	481	13.5
教育	39	4.1	441	12.3
ニュース	30	3.2	258	7.2
買い物	130	13.8	151	4.2
健康	57	6.0	134	3.7
金融	28	3.0	128	3.6
娯楽	40	4.2	85	2.4
Web ホスティング	4	0.4	77	2.2
未分類	60	6.4	76	2.1
疑わしい	3	0.3	76	2.1
政府/法務	5	0.5	73	2.0
旅行	41	4.3	72	2.0
参考	4	0.4	66	1.8
ゲーム	5	0.5	59	1.7
オフィス/ビジネスアプリケーション	1	0.1	51	1.4
Web 広告/分析	4	0.4	51	1.4

表 9 に示す。国内外ともに上位 2 つはテクノロジー/インターネット、ビジネス/経済であった。上位 2 分野で、全体の 35%(国内) と 37.5%(国外) を占めている。国内に限らず、技術志向のこれらの分野は、セッションリプレイを初めとする最新技術の導入に対して積極的である。一方、買い物と旅行のカテゴリは、海外と比べて国内における導入率が高い分野であり、2 倍から 3 倍の導入率を観測している。

6. 導入サイトのプライバシーポリシーにおける公表状況

6.1 対象

5 章で調査した導入サイトのうち、主要 14 セッション

リプレイサービスの種類ごとに、ランダムにサンプリングした 50 サイト、計 400 サイトを調査対象とした。ただし、発見した導入サイトが 50 件に満たないサービスは、発見したサイトのみ調査を行う。また、導入サイトのうち、同一の事業者が運営するものは 1 つのみ調査を行う。

6.2 方法

2023 年 11 月に、全プライバシーポリシーを筆者の一人(梶間)が読み、各項目の有無を判断し、手動で調査を行った。以下の 2 点の記載の有無を調査項目とした。

(1) 利用目的

情報の利用目的にサービスの改善等の記載があること

(2) 導入サービス名の記載

導入を行っている具体的なセッションリプレイサービス名の記載があること

6.3 結果

プライバシーポリシーにおけるセッションリプレイサービスの公表状況について、国内外の調査結果を表 10 に示す。

具体的なサービス名については、国内では 87.7% (300 - 244 = 56 件)、国外では 91% (400 - 36 = 364 件) が公表を行なっていなかった。サービス名については国内外で差はなく、共通して約 9 割が非公開である。

目的については、国内では 18.6% (300 - 244 = 56 件)、国外では 4% (400 - 384 = 16 件) が公表を行なっていなかった。ただし、国外サイトの平均公表率 96% に対して、国内サイトは 81% に留まっており、15% 程低い。(ただし、調査時期が約 1 年間差があることを注意されたし。)

プライバシーポリシーの例を目的のサービス名の記載の有無別に図 3 (目的とサービス名記載) [26] 図 4 (目的のみ記載, サービス名なし) [27] 図 5 (目的もサービス名も記載なし) [28] に示す。目的とサービス名の公表を行っていないウェブサイトの特徴として、ウェブサイト利用時の記載がないものが見られた。

7. おわりに

本研究では、セッションリプレイサービスに対するユーザの許容度の調査とマウスの移動履歴による個人識別性、国外サイトにおけるセッションリプレイサービスの導入、公表状況について調査を行った。

セッションリプレイサービスに対するユーザ感情の調査では、特定の要素ではなく情報の利用目的によってユーザの許容度が変わることを明らかにした。また 62.3% の人がマウスやキーボード操作を取得されることに対して否定的であることがわかった。

マウス操作による個人識別性では、15 人の参加者からマウスの移動履歴の取得を行い識別実験を行った。実験の結

表 10 国内外の導入サイトにおけるプライバシーポリシーでの公表状況

導入サービス	国内 [1]					国外				
	調査数	目的		サービス名		調査数	目的		サービス名	
		件数	%	件数	%		件数	%	件数	%
Microsoft Clarity	50	47	94	12	24	50	47	94	3	6
Hotjar	50	47	94	6	12	50	49	98	6	12
mouseflow	50	31	62	2	4	50	47	94	8	16
crazyegg	50	41	82	10	20	50	50	100	6	12
ContentSquare	27	24	89	2	7	50	49	98	4	8
lucky orange	20	14	70	0	0	27	27	100	0	0
fullstory	5	4	80	1	20	50	48	96	6	12
Yandex	15	11	73	1	7	7	7	100	2	29
Dynatrace	8	8	100	0	0	32	29	91	1	3
Glassbox	4	4	100	2	50	8	7	88	0	0
Smartlook	14	7	50	0	0	13	11	84	0	0
Foresee	2	2	100	0	0	0	0	0	0	0
Inspectlet	3	3	100	0	0	7	7	100	0	0
LogRocket	2	1	50	1	50	6	6	100	0	0
合計	300	244	81.3	37	12.3	400	384	96	36	9

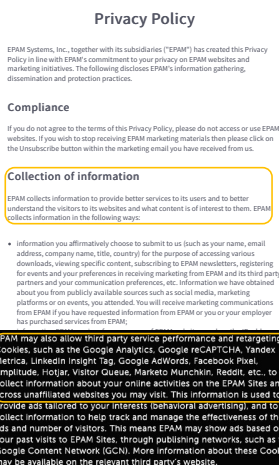


図 3 目的とサービス名の記載のあるプライバシーポリシー（一部抜粋）[26]

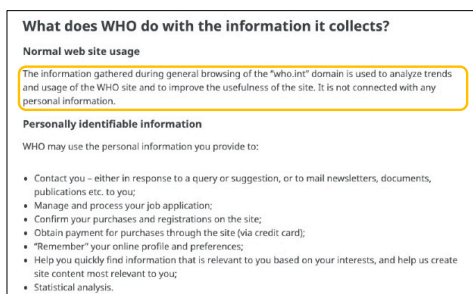


図 4 目的のみ記載のあるプライバシーポリシー（一部抜粋）[27]

果、ランダムより 4 倍高い 0.27 の精度で識別が可能なることを示した。

国外サイトにおけるセッションリプレイサービスの導入、公表状況の調査では、20.1% (3575 件/17,852 件) が導入を行っており、テクノロジー/インターネットやビジネス/経済といったカテゴリのサイトで導入が進んでいることを

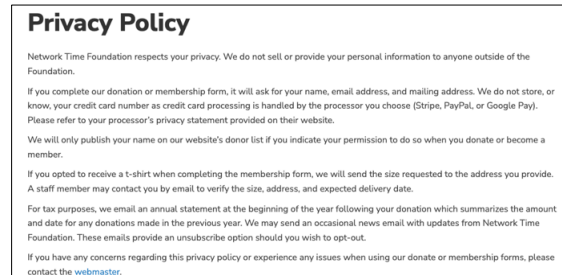


図 5 目的とサービス名の記載のないプライバシーポリシー [28]

明らかにした。また公表状況は、目的が 96% (384 件/400 件)、サービス名が 9% (36 件/400 件) で公表されていることを明らかにした。目的の公表は国内と比較して、国外の方が進んでいることを明らかにした。

以上の調査結果、実験結果を総括して、セッションリプレイサービスに対して利用者の追跡が可能であり、利用者の多くは潜在的な不安を有しているにも拘らず、プライバシーポリシーに依るサービスの利用目的の公表は不十分であることが明らかになった。利用者にセッション取得の事実を分かりやすく伝えるための研究開発を今後の課題とする。

参考文献

[1] 梶間大地, 菊池浩明, “プライバシーポリシーへの記載がない隠しセッションリプレイサービスの導入状況調査”, 第 100 回 CSEC 研究発表会, pp1-5, 2023.
 [2] Gunes Acar, Steven Englehardt, Arvind Narayanan, “No boundaries: data exfiltration by third parties embedded on web pages”, Proceeding of the 20th Privacy Enhancing Technologies Symposium (PETS), pp.220-238, 2020.
 [3] Xiufen Yu, Nayanamana Samarasinghe, Mohammad Mannan, Amr Youseef, “Got Sick and Tracked: Privacy Analysis of Hospital Websites”, IEEE European Symposium on Security and Privacy Workshops (EuroS PW),

- pp.278-286, 2022.
- [4] 泉正夫, 長尾若, 宮本貴朗, 福永邦夫, “マウス操作の特徴を用いた個人識別システム”, 電子情報通信学会論文誌 B, pp.305-308, 2004.
 - [5] 須田 恭平, 石田 繁巳, 稲村 浩, 中村 嘉隆, “日常的な家電操作による人物識別のためのマウス操作による検討”, 情報処理学会研究報告, モバイルコンピューティングと新社会システム研究会, pp.1-6, 2021.
 - [6] 木村悠生, 猪俣敦夫, 上原哲太郎, “深層学習を用いたキーボード入力とマウス操作情報による個人識別”, コンピュータセキュリティシンポジウム (CSS2022), pp.493-499, 2022.
 - [7] 渡邊裕司, 市川俊太, “スマートフォンにおけるタッチ操作の特徴を用いた継続的な個人識別システムの検討”, コンピュータセキュリティシンポジウム (CSS2012), pp.797-804, 2012.
 - [8] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczycki, Wouter Joosen, “Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation”, Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS 2019), pp1-15, 2019.
 - [9] Microsoft Clarity, (<https://clarity.microsoft.com/>, 2023 年 2 月参照).
 - [10] Hotjar, (<https://www.hotjar.com/>, 2023 年 2 月参照).
 - [11] Mouseflow, (<https://mouseflow-jp.com/>, 2023 年 2 月参照).
 - [12] Crazyegg, (<https://www.crazyegg.com/>, 2023 年 2 月参照).
 - [13] Contentsquare, (<https://contentsquare.com/jp-jp/>, 2023 年 2 月参照).
 - [14] lucky orange, (<https://www.luckyorange.com/>, 2023 年 2 月参照).
 - [15] fullstory, (<https://www.fullstory.com/>, 2023 年 2 月参照).
 - [16] Yandex, (<https://metrica.yandex.com/about?>, 2023 年 2 月参照).
 - [17] Dynatrace, (<https://www.dynatrace.com/ja/>, 2023 年 2 月参照).
 - [18] Glassbox, (<https://www.glassbox.com/>, 2023 年 2 月参照).
 - [19] Smartlook, (<https://www.smartlook.com/>, 2023 年 2 月参照).
 - [20] Foresee, (<https://www.verint.com/>, 2023 年 2 月参照).
 - [21] Inspectlet, (<https://www.inspectlet.com/>, 2023 年 2 月参照).
 - [22] LogRocket, (<https://logrocket.com/>, 2023 年 2 月参照).
 - [23] Selenium, (<https://www.selenium.dev/ja/documentation/>, 2024 年 1 月参照).
 - [24] Google Analytics, (<https://marketingplatform.google.com/intl/ja/about/analytics/>, 2023 年 2 月参照).
 - [25] Norton Safe Web, (<https://safeweb.norton.com/>, 2024 年 1 月参照).
 - [26] EPAM — Software Engineering & Product Development Services, (<https://www.epam.com/>, 2024 年 2 月参照).
 - [27] World Health Organization, (<https://www.who.int/>, 2024 年 2 月参照).
 - [28] NETWORK TIME FOUNDATION, (<https://www.nwtime.org>, 2024 年 2 月参照).

付 録

A.1 クラウドソーシングで利用したアンケート

- (1) 参加者個人属性
 - (a) 年代を教えてください
 - (b) 性別についてお聞かせください
 - (c) 普段ウェブサイトを閲覧する頻度を教えてください
 - (d) ウェブサイトを見るときに利用するデバイスを教えてください
 - (e) あなたの IT スキルを教えてください
- (2) 架空サービスに対する許容度 (気にしない, 気持ち悪いが利用する, 気持ち悪く感じ利用しないから選択)
 - (a) あなたのスクロール操作が全て記録されるサービス
 - (b) あなたのマウスの動きから悩んでいる商品を推測し, 類似したおすすめの商品を推薦するショッピングサイト
 - (c) あなたの閲覧時間が記録されているウェブサイト
 - (d) あなたのキーボード入力内容から最適な広告を配信するウェブサイト
 - (e) あなたのマウスの動きからページのデザインを改善しているウェブサイト
- (3) 参加者のセキュリティ意識
 - (a) 普段あなたは初めて訪れたウェブサイトで Cookie を受け入れますか
 - (b) あなたは普段パスワードマネージャを利用しますか
 - (c) この質問には「はい」を回答してください
 - (d) あなたは普段ウェブサイトを利用するとき, プライバシーポリシーを確認しますか
- (4) ウェブサイトからのセッションリプレイサービスの有無の知覚可能性
 - (a) あなたはセッションリプレイサービスを知っていましたか
 - (i) 知っている
 - (ii) 名称は知っているがサービスの概要は知らない
 - (iii) 名称は知らないがそのようなサービスがあることは知っている
 - (iv) 知らない
 - (b) マウスやキーボード操作が取得されることに対してどのように感じるかお聞かせください
 - (i) 取られたくない
 - (ii) あまり取られたくない
 - (iii) どちらでもない
 - (iv) あまり気にしない
 - (v) 気にしない
 - (c) 下記の 3 つのウェブページのうち, 1 つのサイトではセッションリプレイサービスを利用しています. このうち, どのサイトで利用しているか教えてください
 - (d) あなたは Google Analytics というサービスを知っていますか