

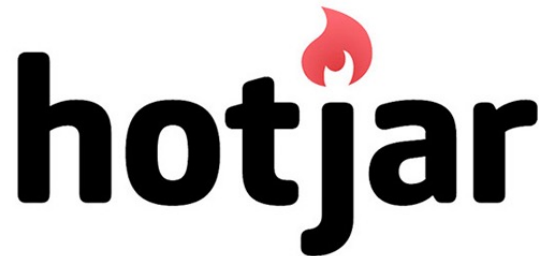
Failure of Privacy Policy for Session Replay Services used for Monitor Your Keystroke

Daichi Kajima and Hiroaki Kikuchi
Meiji University

NBiS 2023

Background

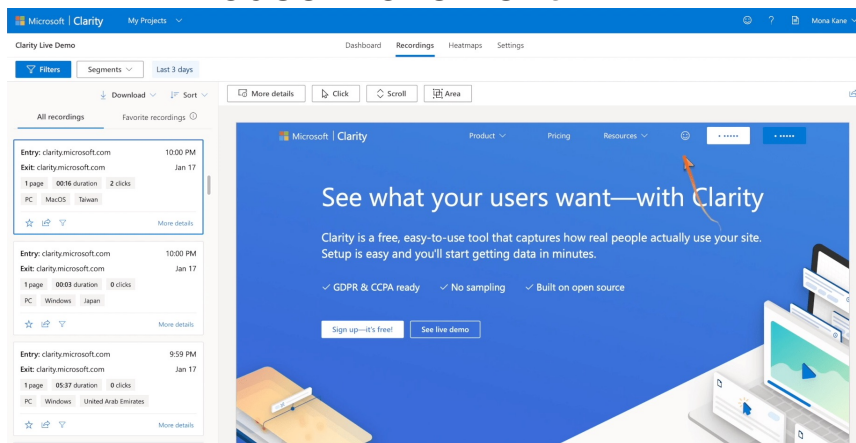
- What is **Session Replay Service**?
 - Capture a visitor's journey on a web site.
 - Improving user experience and identifying obstacle in browsing sites.

The logo for Hotjar, featuring the word "hotjar" in a bold, black, sans-serif font. A red flame icon is positioned above the letter "j".The logo for Crazy Egg, featuring a green egg-shaped icon with a white "C" inside, followed by the text "crazyegg" in a green, lowercase, sans-serif font.The logo for Clarity, featuring a blue and teal geometric icon on the left and the word "Clarity" in a blue, sans-serif font on the right.The logo for Mouseflow, featuring a blue atomic symbol icon above the word "mouseflow" in a blue, lowercase, sans-serif font.

Visitor's journey

- Journey = history of events
 - Mouse movements
 - Regarded as personally identifiable information (pii).

Mouse movement



Heat map



Sensitive information

■ Keystrokes

- ❑ User event captured in session
- ❑ May contain passwords

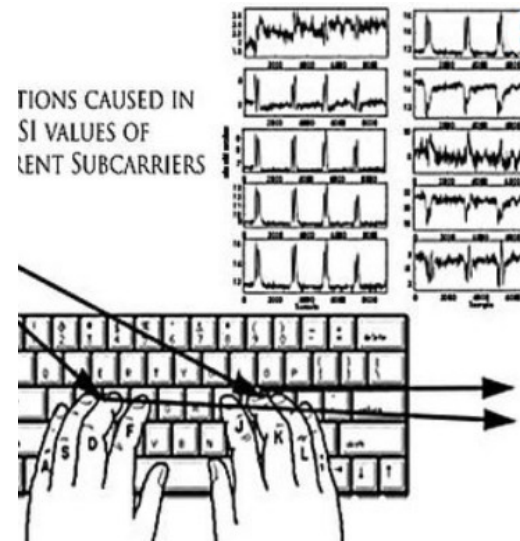
```
▼ 294: {Kind: 18, Args: [762, "k", true, true], When: 4602}
  ▶ Args: [762, "k", true, true]
    Kind: 18
    When: 4602
▼ 295: {Kind: 18, Args: [762, "か", true, true], When: 4797}
  ▶ Args: [762, "か", true, true]
    Kind: 18
    When: 4797
▼ 296: {Kind: 18, Args: [762, "かj", true, true], When: 4848}
  ▶ Args: [762, "かj", true, true]
    Kind: 18
    When: 4848
▼ 297: {Kind: 18, Args: [762, "かじ", true, true], When: 4921}
  ▶ Args: [762, "かじ", true, true]
    Kind: 18
    When: 4921
▼ 298: {Kind: 18, Args: [762, "かじm", true, true], When: 5133}
  ▶ Args: [762, "かじm", true, true]
    Kind: 18
    When: 5133
▼ 299: {Kind: 18, Args: [762, "梶間", true, true], When: 5147}
  ▶ Args: [762, "梶間", true, true]
    Kind: 18
    When: 5147
```

Privacy Regulation

- Act on protection of personal information (APPI), Japan,
 - Article 15: business shall specify the **purpose** of utilizing the personal information as explicitly as possible.
 - » **Privacy policy**
 - Article 16: business shall **not** handle personal information without obtaining in advance a principal's **consent**.
 - » **User consent**.
- GDPR and CPPA

Privacy Breach?

- A website
 - Captures keystroke without exposing session-replay service in the **privacy policy**
- Keystroke
 - Sensitive information can be used as **user authentication**
- **Violation** of the privacy act
 - It could be considered as privacy breach and should notify incident.



Hidden Tracking

- 56 websites (19%)
 - Tracks your journey without your consent
 - No specifying session-replay in their privacy policy

A website using session-replay



What Is TORANOANA?

"Comic Toranoana" is the first leading doujin shop opened in Akihabara Tokyo in 1994 that mainly deals with "Doujinshi" in the industry. Renowned as the first pioneer in the domestic doujin mail order service, and the world's biggest otaku shop.

There are over 100,000 circles of creators registered in Toranoana. In particular, the number of works for women is top-class in the industry, and more than 50,000 works are always available. Toranoana deals with the numerous works of

The privacy policy

とらのあなTOP / 個人情報保護方針

個人情報保護方針

基本方針

株式会社虎の穴（以下、「当社」といいます。）は、通信販売やポイントカードのご利用、商品のご予約申込み、アンケートへのご回答などに際して、お客様の個人情報をお預かりする場合がございます。また、業務上、お取引先やサードパーティ様などのご担当者の個人情報が必要とし、それらをお預かりする場合がございます。当社は、皆様からお預かりした個人情報、お客様にサービスをご提供する上で必要不可欠なものであるとともに、厳密にとつて管理・保護することをご認識しております。当社は、個人情報を取り扱う企業として、個人情報保護法をはじめ関係諸法令を遵守し、情報システムの安全性と信頼性の確保に万全を期し、管理責任者設置のもと、お客様の個人情報を厳密に保管・管理することを最優先事項といたします。

個別方針

◇ 1.個人情報保護への取組み

当社は、個人情報保護を組織的に推進するため、次のとおり、必要な体制やルールを整備し、管理活動を行っております。

1.1.) 法令などの遵守

当社は、個人情報を取り扱う企業としての法的責任や社会的責任を自覚し、これを果たします。当社は、当社のすべての事業活動において、「個人情報の保護に関する法律」（以下、「個人情報保護法」といいます。）およびその他の法令などが遵守されるよう、当社の事業に該当するすべての者に対してこれを周知徹底します。

1.2.) 組織体制

当社は、全社および業務単位で個人情報の取扱いに関する管理者を設置します。また、個人情報に関する教育・啓発、お客様からのお問合せへの対応、監査などについても、それぞれ責任を担い、必要な業務を行います。

1.3.) セキュリティの確保

当社は、個人情報に関して不意のない環境を作るため、セキュリティの確保に努めます。当社は、個人情報への不正アクセスや、お客様の個人情報の紛失、破壊、改ざん、漏えいなどのリスクに対処するため、体制やルールを整備し、必要な施策を講じます。

1.4.) 教育・啓発の実施

当社は、当社の事業に該当するすべての者に対して、教育・啓発を行い、個人情報保護に対する一人一人の意識を高めます。これにより、当社のすべての事業において個人情報の適切な取扱いを徹底してまいります。

1.5.) 監査の実施

当社は、個人情報保護活動の一環として内部監査を実施します。これにより、個人情報保護に関する管理・運営を定期的に見直し、改善してまいります。

◇ 2.個人情報の適切な取扱い

当社は、個人情報をご本人の意に反して取り扱われることがないよう、次の原則に基づいて個人情報を取り扱います。

Challenges

- How to detect session-replay service used in websites?
- How to investigate inconsistency of the privacy policy and their deployment of session replay?

Related Works

- Gunes [2]
 - OpenWPM based investigation of third-party script in 50,000 websites.
 - Security credentials stored in a DOM.
- Xiufen [3]
 - Detected 690 sites having session-replay service out of 19,483 hospital websites.

Questions

- How common websites do set up session-replay service?
- How frequent are privacy policies explicitly for specifying a use of session-replay?
- Which section of business are the most explicit for disclosing use of session-replay service?



Our methodologies

- 1. Detection
 - **PublicWWW**, a search engine for HTML source codes, is used to detect **top-15 service** URLs from **top 11,523 websites** in Japan.
- 2. Policy survey
 - **Manual investigation** of **197 sampled** websites chosen from each of TOPIX-17 (major business sectors in Japan).

Result 1: deployments

Srvce	N
Microsoft Clarity	702
Hotjar	89
Mouseflow	68
Yandex	4
ContentSquare	20
CrazyEgg	40
Dynatrace	2
foresee	1
fullstory	6
glassbox	2
inspectlet	0
logrocket	0
luckyorange	7
Smartlook	2
Total	943

Total of 943 websites deployed session-replay service.

It accounts for **8.2%** of 11,523 major websites in Japan.

Result 2: Business sectors

sectors	Clarity	Hotjar	MR	Yandex	CS	CE	DT	FS	fullstory	GB	IL	LR	LR	SL	total
ICT	274	33	22	3	2	6	0	0	3	1	0	0	1	1	346
wholesale	127	19	9	0	11	19	0	1	1	0	0	0	4	1	192
industry	40	13	5	0	4	6	2	0	1	0	0	0	1	0	72
hotel	21	6	1	0	1	0	0	0	0	1	0	0	0	0	30
educational	18	5	4	0	0	1	0	0	1	0	0	0	0	0	29
entertainment	20	3	2	0	0	0	0	0	0	0	0	0	0	0	25
academic study	34	2	3	0	0	1	0	0	0	0	0	0	0	0	40
general service	76	3	8	0	0	3	0	0	0	0	0	0	1	0	91
financial business	13	1	3	0	2	1	0	0	0	0	0	0	0	0	20
construction	7	0	3	0	0	0	0	0	0	0	0	0	0	0	10
electric power	4	0	1	0	0	0	0	0	0	0	0	0	0	0	5
medical service	6	2	1	0	0	0	0	0	0	0	0	0	0	0	9
real estate	15	0	5	1	0	0	0	0	0	0	0	0	0	0	21
transportation	7	0	1	0	0	0	0	0	0	0	0	0	0	0	8
integration	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
unknown	39	2	0	0	0	3	0	0	0	0	0	0	0	0	44
total	702	89	68	4	20	40	2	1	6	2	0	0	7	2	943

ICT: 346/943

(37%)

Wholesale:

192/943

(20%)

Results 3: Policy coverages

Service	N	Service name	Objectives	Neither
Microsoft Clarity	50	12	35	3
Hotjar	50	6	41	3
Mouseflow	50	2	29	19
CrazyEgg	50	10	31	9
ContentSquare	26	2	22	3
Luckyorange	20	0	14	6
fullstory	6	1	2	1
Yandex	15	1	10	4
Dynatrace	8	0	8	0
glassbox	4	2	1	0
Smartlook	14	0	7	7
Foresee	2	0	2	0
Inspectlet	3	0	3	0
logrocket	2	1	0	1
total	300	37	207	56

12.3%

37 websites specify both session-replay service name and objectives of session information.

18.7%

57 websites do not disclose about session-replay at all.

Detected websites

The screenshot shows the Yomiuri Newspaper website. The main article is titled "偽のロイター記者、中国の民主活動家に接触...ゼロコロナ抗議グループの情報探る" (Fake Reuters reporter, contact with Chinese democracy activists... information on Zero-COVID protest groups). The article text mentions a fake Reuters reporter who contacted Chinese democracy activists in February 2022. The sidebar features a grid of product recommendations with prices, such as a 1.780円 bottle, a 2.170円 bottle, a 3.740円 bottle, and a 2.170円 bottle.

Yomiuri Newspaper
<https://www.yomiuri.co.jp/>



Explicitly specifies
 the session-replay
 service in policy

The screenshot shows the Toranoana website, a doujinshi site. The main content is a "NIJISANJI ARCHIVES" page for the year 2022, with a date of 2023年3月7日発表. The page features a grid of product listings with prices, such as a 1.780円 bottle, a 2.170円 bottle, a 3.740円 bottle, and a 2.170円 bottle. The sidebar includes a search bar and a list of categories like "電子書籍" (Digital books) and "電子書籍2冊内" (Digital books 2 volumes).

Toranoana
 (Doujinshi sites)
<https://www.toranoana.jp/>



Hidden tracking
 without user
 consent

Google Analytics

- Comparison with session replay service
 - 116 websites specify the deployment of Google Analytics in their privacy policies.
 - Higher fraction than that of session-replay.

	Session replay service	Google Analytics
deploy	300	277
disclosure	37 (12.3%)	116 (41.6%)

Discussion

- Why so many sites fail to specify session-replay service?
 - Many business operators did not notice that user events are regarded as pii.
 - Most users do not read a privacy policy carefully and do not know at all.
 - The privacy commissioners may not notice that it can violate the regulation.

Mitigation

- 1. **Service providers** should disclose the list of acquired personally identifiable information from the web site in their privacy policy explicitly.
- 2. **Service providers** should obtain visitor's consent for acquirement of session related information including mouse movements and keystrokes.
- 3. **Users** should notice the privacy policy statement before they visit web sites.
- 4. **Users** should avoid unnecessary providing personally unidentifiable information through web site

Conclusions

- We have studied the current risk of session-replay service for violating privacy regulations via capturing visitor's mouse-movement and keystroke without obtaining consent.
- Our analysis find that 981 (8.51%) websites deploying session-replay without specifying their policy.
- We plan to investigate international statistics related to session replay.

