Server Privacy Control: a server-to-client privacy opt-out preference signal

Don Marti
Raptive
dmarti@raptive.com

Introduction

The most enduring and valuable uses of the Internet come from people who choose to communicate, tell stories and socialize online on their own sites, on their own terms, where they are the moderator and control what they choose to share and promote. While each individual blog or other personal site might be a labor of love for one person, collectively the independent web is a major source of educational and cultural wealth, as well as a significant commercial enterprise. Raptive, one of several providers of services to independent sites, is ranked among the top ten media companies by Comscore.

Recently, a key set of concerns for the independent content creators supported by Raptive has been AI crawling of their sites and use of their content. Some of these concerns are based on competition and search considerations, but a more complex set of issues include AI misrepresentation, “hallucinations,” deepfakes and other false content that AI can create based on people’s sometimes deeply personal sites. Solove and Hartzog write, “Privacy law regularly protects publicly available data, and privacy principles are implicated even when personal data is accessible to others.” [SCRAPING]

A person who chooses to communicate online using a Big Tech platform can often take advantage of AI training preferences specific to that locked-down environment, but on the open web, protection is less assured. The centralization and competition concerns are real. We believe that an individual who chooses to express themselves in their own online space should not be placed at a disadvantage relative to a user of a centralized service. The right to control the use of your own words and images by AI is more than a copyright or other commercial right; it is a human right.

Proposed solution

Taking inspiration from the opt-out preference signals (OOPSs) that can pass a user's privacy preferences from a client—a web browser—to a server, Raptive is developing a server-side OOPS that works the same way, but for communicating a preference from a server to a client—a crawler or scraper.

Server Privacy Control (SPC) is a new value for the X-Robots-Tag HTTP header and the meta name=robots meta tag. Here are examples showing SPC alone and with the "noai" and "noimageai" values.

```
X-Robots-Tag: SPC

X-Robots-Tag: noai, noimageai, SPC


<meta name="robots" content="SPC">

<meta name="robots" content="noai, noimageai, SPC">
```

The meaning of SPC is the same as the meaning of Global Privacy Control, but with client and server exchanged. The GPC spec states [GPC]:

> A do-not-sell-or-share preference is when a person requests that their data "not be sold or shared" for instance by activating a Global Privacy Control setting with their user agent or by using tools that default to such a setting (possibly because this setting matches the most common expectations of that tool's users). When set, this preference indicates that the person expects to browse the Web with do-not-sell-or-share interactions.

The equivalent for SPC (differences in bold) is:

> A do-not-sell-or-share preference is when a person requests that their data "not be sold or shared" for instance by activating a **Server Privacy Control** setting with their **web server software** or by using **web server software** that defaults to such a setting (possibly because this setting matches the most common expectations of that tool's

> users). When set, this preference indicates that the person expects to **create content for** the Web with do-not-sell-or-share interactions.

SPC is simply the same opt-out preference signal as GPC, and intended to be interpreted and enforced in the same way. The only difference is that it is passed from a server to a client. SPC is not intended to define or send any new opt-out preference not already expressed by GPC.

SPC as a lightweight approach to AI privacy compliance

AI crawler operators have a strong incentive to implement SPC, because it is a simpler path to achieving independent web site creators' protection goals than other options available under privacy laws and regulations. The trend in privacy law is for greater transparency and explanation of decision-making, which is, for AI, a difficult research area.

Because of unsolved research questions in the area of AI explainability, operators of AI service face challenges in complying with privacy laws that require some parties to disclose personal information to people. For example, under California law, a "business" is required to disclose "inferences" about an individual on request. [INFERENCES]

> [F]or purposes of responding to a request to know, it does not matter whether the business gathered the information from the consumer, found the information in public repositories, bought the information from a broker, inferred the information through some proprietary process of the business's own invention, or any combination thereof. If the business holds personal information about a consumer, the business must disclose it to the consumer on request. We emphasize that, once a business has made an inference about a consumer, the inference becomes personal information—one more item in the bundle of information that can be bought, sold, traded, and exploited beyond the consumer's power of control. Accordingly, inferences satisfy the first condition of the "personal information" inquiry regardless of whether they have been generated internally by the responding business or received from another source.

By respecting SPC, and giving the operator of a personal site the assurance that personal information on the site is not being scraped for training AI to form inferences about people, the firm can keep personal information, from which inferences might have been drawn, out of

training sets in the first place and avoid downstream problems. Training on personal information must be avoided in order to minimize issues with “right to know” and “right to delete” that they would otherwise receive from people concerned about AI inferences formed from their own information and from information about other people such as family members mentioned on their site. A commenter or forum participant on an independent site might also share personal information with the expectation that it would be processed in accordance with the site’s own privacy policy; applying SPC as appropriate would also enable a site to meet the needs of these users.

Next steps

OOPSs are not just a necessary time-saver for Internet users in opt-out-based privacy jurisdictions such as California, but also a practical improvement in signaling for other jurisdictions [GPC-GDPR]. We plan to pursue registration of SPC as a OOPS in jurisdictions such as Colorado where registration is required and look forward to collaborating with the AI development community.

References

[GPC]
Zimmeck, Sebastian et al. 2024. Global Privacy Control Specification Proposal.
URL: https://privacycg.github.io/gpc-spec/

[GPC-GDPR]
Berjon, Robin. 2021. GPC under the GDPR.
URL: https://berjon.com/gpc-under-the-gdpr/

[INFERENCES]
Bonta, Rob, and Lee, Susan Duncan. 2022. Opinion No. 20-203.
URL: https://oag.ca.gov/system/files/opinions/pdfs/20-303.pdf

[SCRAPING]
Solove, Daniel J. and Hartzog, Woodrow, The Great Scrape: The Clash Between Scraping and

Privacy (July 03, 2024). Available at SSRN: https://ssrn.com/abstract=4884485 or http://dx.doi.org/10.2139/ssrn.4884485