Quantifying Digital Footprints: a Client-Side Risk Scoring and Visualization

Vasudha Phaltankar, Vidya Harkal, Sankalp Mali*, Sarvesh Kuvalekar, Aniket Jaitkar, and Tanmay Malkar

Department of Computer Engineering,
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India
vasudhaphaltankar@gmail.com, vidya.harkal90@gmail.com,
malisankalp2003@gmail.com*, sarveshkuvalekar2207@gmail.com,
aniketjaitkar1432@gmail.com, malkartanmay6300@gmail.com

Abstract. Digital footprints are the large amounts of user-specific digital traces that have built up because so many people use online services and third-party analytics. Companies can follow users across sites and make behavioural profiles using these traces, which include cookies, persistent web storage artifacts, network requests, and browser-level finger-printing signals. Most of the time, users don't even know they're being tracked [3], [8]. Most current privacy tools only stop one type of tracking at a time, and they don't do a good job of letting you see how much tracking exposure you get on different websites [9]. This paper intro-duces a client-side quantitative framework for evaluating web tracking exposure via the comprehensive analysis of cookies, storage artifacts, and third party network interactions. We suggest a clear, rule-based privacy risk scoring model that takes into account things like identifier persistence, third-party prevalence, and security attributes. Comparing how users track their behavior across various websites is made possible by the framework, which transforms low-level browser telemetry into intelligible measurements and visualisations. An empirical analysis conducted on a diverse collection of real-world websites shows notable differences in tracking intensity between browsing categories, supporting the effectiveness of the suggested metrics in distinguishing between high- and low-risk environments. Since all analysis is completed within the user's browser, no data must be transferred to other locations, protecting their privacy. The proposed method focuses on measurement and interpretability instead of blocking trackers. This gives users and researchers useful information about their digital footprint exposure while still protecting their privacy.

Keywords: Digital Footprint · Privacy Protection · Web Tracking · Data Visualization · Cookies · Browser Extension · User Analytics · Risk Assessment · Data Normalization · Fingerprinting Detection · Sankey Diagrams · Web Security · Online Privacy · Consent-Based Data Collection

© The Author(s) 2026
M. Kaushik et al. (eds.), *Proceedings of the International Conference on Responsible, Risk-aware, and Regulated AI (RRRAI 2026)*, Advances in Intelligent Systems Research 210,
https://doi.org/10.2991/978-94-6239-723-1_41

1 Introduction

The modern web has become a huge network for sharing information because of the rapid growth of online services, personalized marketing platforms, and third-party analytics ecosystems. Every time someone interacts with a website, like by visiting a page, filling out a form, clicking a link, or loading multimedia content, they leave behind digital traces that can be seen, such as HTTP cookies, local and session storage artifacts, network requests, and script-level fingerprinting signals [3], [8]. All of these things make up a user's digital footprint, which lets you make detailed profiles of how they browse different websites.

While these mechanisms enable usability, personalization, and optimization, there has also been a rising concern related to their impact on user and entity behavior analysis, which may result in unauthorized profiling, targeted advertising, inference of identities, and behavior correlation without users' consent because these mechanisms and frameworks enable easy and efficient tracking using persistent identities and crossbrand tracking systems [3], [4], [8]. Recently, research has also emerged indicating that it has become more challenging for users to understand how their traces have been collected, shared, and reused on different platforms because of the rising complexity involved in digital traces [9], [10]. While there have been some improvements made through existing privacy solutions on the browser and tracking-resistance plugins, they remain concentrated on combating individual tracking entities, such as cookies and third-party scripts, without addressing the bigger picture of cumulative tracking risk awareness [3], [8]. This results in the end user having little understanding of how their accumulated tracking risk is made up of several tracking aspects, including cookies, storage, and network communications [9].

Additionally, most of the analytical methods utilize centralized or server-based processing, which also brings a different set of risks associated with Minimization of Data and Compliance with laws [3].

Recent interdisciplinary studies stress the importance of transparent, measurational approaches to identifying the extent to which traces of identity and behavior are laid bare on the Internet [9], [10]. Visualization-assisted analyses have been observed to increase the comprehension of complex data streams and relations, especially if abstract telemetry is converted into simple visual forms such as flow diagrams and timelines [2], [6]. However, the dominant visualization-assisted software tools are currently descriptive and have no quantification components assessing risks on different websites or types of browsing.

In this paper, we present a client-side quantitative analytical framework for quantifying web tracking exposure to overcome these limitations. It unifies into one measurement pipeline for cookies metadata, persistent storage artifacts, and third-party network requests. A rule-based privacy risk model based on an open-scoring model quantifies exposures with regard to aspects such as identifier persistence, third-party prevalence, and security attributes and allows for systematic comparison between websites. The derived measurements are then conveyed through interactive visualizations that allow better interpretability and user awareness.

Concretely, the main contribution of this work is not the blocking of trackers themselves, but rather the quantification of tracking exposure using an interpretable in-browser analysis framework. A empirical evaluation of said metrics across a diverse corpus of real-world websites shows significant variance in tracking intensity among categories, validating that the proposed metrics successfully distinguish between high- and low-risk sites. All analysis is executed on the browser client-side, per the privacy-by-design principle, eliminating the need to transmit data to third-party services.

The rest of this paper is organized as follows: In Section II, related work on digital footprint analysis and visualization of privacy is briefly introduced from papers [1], [9], and [10]; The system architecture and data model will be proposed in Section III; The risk scoring method and visualization solution will be explained in detail in Section IV; The experimental results and comparisons will be given in Section V; Finally, limitations and future research directions will be explored in Section VI.

2 Literature Survey

2.1 Digital Footprints and Online Tracking

Digital footprints encompass the aggregate of observable data traces produced through user engagement with online platforms, encompassing cookies, persistent storage objects, network requests, and behavioural identifiers. Industry and security analyses underscore that contemporary websites are progressively dependent on persistent client-side identifiers and third-party analytics frameworks to facilitate user profiling and targeted advertising [3], [8]. Palowise [4] and Formica AI [7] additionally emphasise the swift proliferation of tracking technologies and the escalating magnitude of data collection within web ecosystems.

Digital footprint studies have shown how a person's digital footprint is composed of numerous bits of information, and they depict how these online traces can provide sensitive information related to an individual's identity and behavior [9], [10]. Researchers Olinder et al. [10] show that digital footprints are being used more and more frequently in a multitude of social science disciplines, leading to an increased level of concern over transparency of users and their knowledge of how they are tracked. The need for tools that can help to expose and analyze tracking behavior—while making it understandable to end users—was emphasized throughout these findings.

2.2 Privacy Risks and Limitations of Existing Protection Tools

A large number of anti-tracking browser-based privacy tools or extensions are developed to protect against vulnerabilities created by online tracking. The major anti-tracking strategies employ blocking third-party scripts, isolating cookies, or blocking cross-site requests. While these types of privacy tools reduce the ability to track someone through one of these avenues, security evaluations of these

tools frequently show that they generally address only an individual tracking method without factoring in how multiple identifiers work together [3], [8]. Also, the majority of privacy tools provide a preventative approach and do not give users an understanding of the different aspects of tracking and how they can compare their potential privacy risk level from one website to another. According to a study by ResearchGate, one of the biggest obstacles to creating a user's understanding and ability to make decisions is the lack of measurement-based feedback regarding his/her digital footprint visualisation ([9]). As a result, even when using privacy-enhancing tools, many users would still continue to access websites that pose high privacy risks.

2.3 Measurement and Analysis of Digital Footprints

New research shows that measurement-based methods can provide insight into the impact of digital footprints. Instead of focusing on blocking or anonymizing users, measurement approaches will help companies quantify how much they track their users on multiple platforms [9], [10]. Walczak and Marwaha Kraetzig [5] show how to use large-scale digital tracks to understand behaviour and validate the importance of using structured telemetry to analyse data. But there are not many ways to analyse data on a browser without requiring users to have their data processed at a central server, which creates privacy issues and reduces users' trust in the companies involved [3]. Thus, there is little evidence that there are ways to analyse data on the user's machine without sending the data to the server.

2.4 Visualization Techniques for Privacy and Tracking Analysis

In order to convert complicated telemetry data into understandable insights, visualization is essential. Structured visual layouts improve the user's understanding of system behavior, according to previous work on wireframe modeling and dashboard design [1]. Industry research also shows that interactive visualizations, such as timelines, graphs, and flow diagrams, enhance the' ability to spot trends and abnormalities in large datasets [2], [6]. Network graphs and flow-based representations are especially useful for showing third-party data sharing relationships, according to research on digital footprint visualisation [9]. Despite these developments, the majority of visualization-focused tools are still descriptive and do not integrate with quantitative risk models, which limits their usefulness for cross-site comparisons.

2.5 Research Gap and Motivation

It is clear from the reviewed literature that current solutions pay little attention to unified, client-side tracking exposure measurement and instead focus on tracker blocking, descriptive visualization, or centralized analytics. Frameworks that incorporate various tracking vectors, such as cookies, persistent storage, and

network interactions, into a transparent, quantitative risk assessment pipeline that runs fully within the browser are scarce.
The proposed work, which focuses on measuring digital footprint exposure through an interpretable, browser-resident framework that integrates measurement, visualization, and empirical evaluation while upholding privacy-by-design principles, is motivated by this gap.

3 Methodology

3.1 System Synopsis

The high-level architecture of the suggested clientside digital footprint analysis framework is shown in Fig. 1. The system is implemented as a browserresident extension that records tracking-related telemetry by passively observing user browsing activity. As demonstrated, content scripts and browser APIs are used to gather data created during webpage interaction, including cookies, storage objects, and third-party network requests. A rulebased risk analyser and classification module processes these observations locally. After analysis, the results are displayed on interactive dashboards and can be exported as structured reports. No raw browsing data is sent to external servers because the entire pipeline runs inside the browser environment.

3.2 Data Collection

The process for the collection of the data is intended to be transparent when the user is performing a normal browsing session. This is because, when the user visits a website, the system retrieves artifacts regarding the process of page loading and interaction, which entail HTTP cookies, local and session storage, and network requests targeting third-party domains.Content scripts within the extension track Document Object Model (DOM) interactions as well as storage accesses, which are directly observable at the page level. Correspondingly, the browser extension API interface is employed to garner additional metadata, which cannot otherwise be ascertained from the DOM model, such as the attributes of cookies or the origin of the request. The domain name, expiration time, security indicators, and request context of each artefact will then be used to change it.

3.3 Risk Scoring Model

The framework measures tracking exposure by using a clear, rule-based model to score privacy risks. For each tracking artifact that is seen, clear standards are used to rate how it affects privacy. Tracking artifacts are scored and rated based only on the score associated with that artifact's ability to identify its source over time, if it is a first-party or third-party domain, and if it has secure features such as an HTTP or HTTPS protocol. Each tracking artifact has a series of

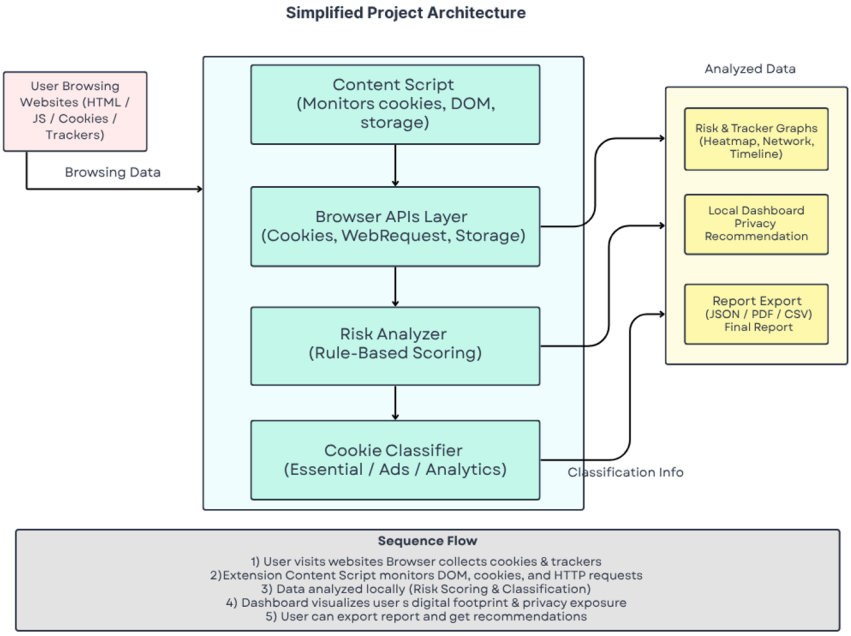

Fig. 1. Simplified System Architecture – Browser-based Digital Footprint Tracker and Visualizer

flags assigned to it which serve as partial scoring elements for each of the three rating criteria. Once all the flags for a tracking artifact are collected against each of the rating criteria, they are then combined to create a single Privacy Risk Score (PRS) for that specific website or browsing session. The Privacy Risk Score can then be adjusted between 0 and 100, with higher numbers indicating that a website or browsing session is more likely to expose users to tracking. This deterministic method for generating Privacy Risk Scores will make it much easier for users and researchers to see how different forms of tracking behavior contribute to a final determination.

3.4 Classification Logic

The cookie/tracker categorization uses three categories; Essential, Analytics, and Advertising. Deterministic rules are used to look at the attributes, name, duration, etc. of each tracker/cookie and of any corresponding domains to identify the behavioural patterns associated with this type of cookie/tracker. Authentication and session control are two examples of important artefacts that are needed for the site to work. The Analytics artefacts work with services that keep track of how well things are working and how often they are used. The advertising artifacts work with behavioral targeting and tracking across sites. Without the use of machine learning models or third-party block lists, these categories enable

users to distinguish between tracking techniques that are essential and those that are probably intrusive.

3.5 Privacy Preservation

This strategy's main goal is to safeguard users' privacy. Everything takes place in the browser, which gathers the data, examines it, and even creates the images. You don't have to be concerned about things escaping. You know, sensitive things like cookie IDs simply don't last forever. Alternatively, if they process it, it is altered and cannot be undone before verifying everything. In this manner, no raw data is sent to servers elsewhere. That seems to keep things safe. Users have local control over exporting and reports. This, in my opinion, satisfies those privacy by design concepts and is also fairly simple. Even so, you can analyse without disclosing a lot of information. It can feel a little difficult to balance that at times.

4 Experimental Evaluation

4.1 Experimental Setup

To see how well the proposed framework works the results were compared to how other privacy extensions do their job. Old tools just stop trackers but they do not give you any idea of how much you are really exposed. The proposed framework is different because it lets you measure and score the tracking that is going on and you can see it in a way that's easy to understand. This makes it easy to compare how different websites and categories are doing when it comes to tracking. This comparative analysis demonstrates how, even in situations where tracker blocking is not an option, measurement-oriented tools may aid in increasing transparency and assisting individuals in making better decisions.

4.2 Evaluation Metrics

Tracking exposure was measured using the following metrics:

- **Tracker Count (TC):** This is the number of websites that are contacted by a website. We call this the Tracker Count. The Tracker Count is, like a list of all the websites that a website talks to.
- **Cookie Persistence (CP):** The mean number of days that cookies last.
- **Insecure Cookie Ratio (ICR):** The percentage of cookies that do not have Secure and/or HttpOnly properties is called the Insecure Cookie Ratio (ICR).
- **Fingerprinting Events (FE):** Detected script level access to browser fingerprinting-related APIs.
- **Privacy Risk Score (PRS):** A normalized score in the range of 0–100 computed from the above metrics.

These metrics were selected to capture both the scale and severity of tracking mechanisms present on visited websites.

4.3 Website-Level Results

The experimental findings indicate significant variability in tracking behavior among websites and categories. Social media sites always had more third-party trackers and cookies that lasted longer, which made privacy risk scores go up.On the other hand, websites that provided information and education showed less tracking activity and less identifier persistence. The representative results for each website category are compiled in **Table 1**. These results show that the suggested metrics successfully distinguish between browsing environments with high and low risk.

Table 1. Category-Wise Tracking Exposure

Website Category	**Avg. Trackers**	**Avg. Cookie Age (days)**	**Avg. Risk Score**
Social Media	High	Long	High
E-commerce	Medium	Medium	Medium
News & Media	Medium	Short–Medium	Medium
Education	Low	Short	Low

4.4 Comparative Analysis

To determine how well the proposed framework could analyse data, we qualitatively compared the outcomes with the behaviour of existing privacy extensions. Conventional tools typically only block or hide trackers; they don't provide you with cumulative exposure numbers. However, by offering explicit measurement, risk scoring, and visualising tracking activity, the proposed framework enables systematic comparison across websites and categories. In addition to tracker blocking, this comparative analysis demonstrates how measurement-oriented tools can improve transparency and well-informed decision-making.

4.5 Discussion

According to the experimental evaluation, a browser-resident client-side framework that eschews external data aggregation can be used to measure and visualise exposure to digital footprints. The validity of the chosen metrics is further supported by observed trends that are consistent with earlier studies on digital footprints. The framework protects user privacy while enabling meaningful evaluation of web-tracking behaviour by combining quantitative risk scoring with easily comprehensible visualisations.

5 Conclusion and Future Work

A client-side quantitative framework is proposed to examine exposure to digital footprints through cookies, persistent storage artifacts, and third-party network interactions. The method focuses on measurement and comprehension, enabling users to ascertain the degree of tracking without depending on ambiguous blocking techniques or centralized aggregation. The framework is based on the idea of "privacy by design," which means that all analysis is done in the browser. Users can easily understand how web tracking works with this.

Testing on a lot of real-world websites showed that tracking exposure changed a lot depending on the type of browsing. This proved that the suggested risk scoring metrics can tell the difference between environments with low and high risk. The results show that browser resident analysis is an effective way to raise user awareness and make it easier to compare online privacy risks.

Future work will explore adaptive methods for uncovering novel fingerprinting techniques, perform longitudinal studies of tracking behaviour over extended browsing sessions, and implement the framework across multiple browsers to improve generalizability. Additional refinement of the scoring model and user studies on how visual feedback affects privacy awareness also hold strong promise.

References

1. Milutinović, J. S., Milić, P. O., Vučević, S. S.: Wireframe modelling of web-based open data applications. Serbian Journal of Applied Mathematics and Informatics **16**(2), 111–123 (2024)
2. Appinventiv: AI-powered data visualization: Unlock dynamic insights in 2025. (2025)
3. NCC Group: Digital footprint review. NCC Group Threat Intelligence (2023)
4. Palowise: 37 digital footprint statistics and trends to know for 2024. (2024)
5. Walczak, S., Marwaha Kraetzig, N.: From digital footprints to urban intelligence: Leveraging public GPS traces. Geoawesome (2024)
6. Zapier: The 6 best wireframe tools in 2025. (2025)
7. Formica AI: Top digital footprint trends of 2023. Formica AI Blog (2023)
8. Bitdefender: The digital you in 2023: Modern digital footprints and their impact on privacy and security. (2023)
9. ResearchGate: Constructing, visualizing, and analyzing a digital footprint. (2023)
10. Olinder, N., Tsvetkov, A., Fedyakin, K., Zaburdaeva, K.: Using digital footprints in social research: An interdisciplinary approach. WISDOM **16**(3), 124–135 (2020)
11. Allen, H. G., Stanton, T. R., Di Pietro, F., Moseley, G. L.: Social media release increases dissemination of original articles in the clinical pain sciences. PLOS One **8**(7), e68914 (2013)
12. Bornmann, L.: Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. Scientometrics **103**(3), 1123–1144 (2014)
13. Bornmann, L.: Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. Journal of Informetrics **8**(4), 895–903 (2014)

14. Bornmann, L., Haunschild, R.: Which people use which scientific papers? An evaluation of data from F1000 and Mendeley. Journal of Informetrics **9**(3), 477–487 (2015)
15. Cheung, M. K.: Altmetrics: Too soon for use in assessment. Nature **494**(7436), 176 (2013)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.