2026 Volume: 02 Issue: 06 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-17

A UNIFIED PRIVACY RISK INDEX FOR QUANTIFYING ECOSYSTEM-LEVEL WEB TRACKING LEAKAGE

***Rohit Kumar**

Department of Computer Science & IT Magadh University, Bodh Gaya, India.

Article Received: 29 April 2026
Article Revised: 19 May 2026
Published on: 09 June 2026

***Corresponding Author: Rohit Kumar**
Department of Computer Science & IT Magadh University, Bodh Gaya, India.

DOI: https://doi-doi.org/101555/ijrpa.5834

ABSTRACT

Existing studies of web tracking primarily detect tracking mechanisms but lack quantitative models capable of estimating cumulative privacy leakage across interconnected tracking ecosystems. This paper introduces a unified Privacy Risk Index (PRI), a multivariate quantitative framework that integrates identifier entropy, tracker density, cross-site behavioral correlation, third-party data-sharing frequency, and network-structural centrality into a single ecosystem-level leakage function. The proposed formulation establishes bounded sensitivity and interpretable parameter influence, enabling stable comparative privacy-risk estimation across heterogeneous tracking environments. The framework is evaluated using controlled simulations and automated web-measurement datasets, and comparative experiments against entropy-based and exposure-score metrics demonstrate that PRI provides improved discrimination of high-risk tracking configurations while capturing nonlinear amplification effects arising from tracker concentration and cross-site correlation. The resulting metric offers a computationally efficient quantitative indicator for ecosystem-scale privacy auditing, mitigation evaluation, and regulatory monitoring of web tracking infrastructures.

KEYWORDS: Privacy Leakage, Web Tracking, Privacy Risk Index, Quantitative Privacy Measurement, Tracker Ecosystems, Network Centrality, Cross-Site Tracking.

1. INTRODUCTION

Web tracking has become a foundational component of the modern digital economy, enabling large-scale collection of user behavior through cookies, third-party scripts, fingerprinting techniques, and real-time bidding infrastructures. These mechanisms allow tracking entities

to aggregate browsing histories, infer personal attributes, and construct persistent user profiles across domains. Despite the existence of regulatory frameworks such as the General Data Protection Regulation (GDPR) and India's Digital Personal Data Protection Act (DPDP 2023), the underlying technical ecosystem remains highly opaque, fragmented, and difficult for users to meaningfully assess.

Most prior research on web tracking has focused on detection rather than quantification. Large-scale measurement studies have documented the prevalence of third-party trackers, fingerprinting scripts, and cross-site identifiers, revealing the dominance of a small number of powerful tracking entities. Other works have examined specific mechanisms such as browser fingerprinting, cross-device linking, and data broker networks. While these studies provide valuable descriptive insights, they offer limited tools for estimating the magnitude of privacy leakage produced by the combined effects of multiple tracking factors.

Privacy leakage is not driven by a single mechanism. It emerges from the interaction of several dimensions: how predictable identifiers are, how many trackers are present, how strongly user behavior is correlated across sites, how frequently data is shared with third parties, and how centrally tracking entities are positioned within the data-exchange network. Existing approaches typically analyze these factors in isolation. As a result, they fail to capture how structural concentration and cross-domain aggregation can amplify privacy risk even when individual signals, such as identifier entropy, appear strong.

Information-theoretic metrics, such as entropy, have been used to evaluate identifier unpredictability, while network-based analyses have highlighted the centralization of tracker ecosystems. However, these approaches are rarely integrated into a unified, quantitative risk model. Moreover, most privacy assessments remain qualitative, descriptive, or binary, offering little support for comparative evaluation across websites, configurations, or regulatory environments.

This lack of a systematic, quantitative metric creates a practical gap for privacy engineers, auditors, and policymakers. Without an interpretable risk index, it is difficult to compare tracking environments, evaluate mitigation strategies, or communicate privacy exposure in measurable terms. Users are left with vague notions of "high" or "low" risk, while regulators lack technical indicators that reflect real-world tracking dynamics.

To address this gap, this paper proposes a mathematical and computational framework for quantifying privacy leakage using a unified Privacy Risk Index (PRI). The PRI integrates five key dimensions of tracking-driven exposure: identifier entropy, tracker density, cross-site behavioral correlation, third-party data-sharing frequency, and network-level leakage derived

from structural centrality. To the best of our knowledge, this is the first study to integrate information-theoretic, behavioral, and network-structural factors into a unified quantitative privacy risk index for web tracking ecosystems. By combining information-theoretic and graph-based concepts, the model captures both local identifier properties and global ecosystem effects.

The framework is evaluated through controlled simulation of diverse tracking environments, allowing systematic analysis of how individual parameters and their interactions influence overall privacy risk. The goal is not to detect trackers, but to quantify how their collective behavior translates into measurable privacy exposure. This study focuses on interpretable ecosystem-level measurement rather than detection of individual tracking mechanisms.

By providing an interpretable and computationally lightweight metric, this work contributes a practical tool for comparative privacy assessment, regulatory auditing, and privacy-aware system design. While the current evaluation is simulation-based, the model is directly extensible to real-world telemetry and measurement datasets.

While our earlier work proposed a mathematical model for quantitative privacy risk estimation in web analytics, the present study is fundamentally different in scope, structure, and objectives. The prior model focused on single-site analytics environments and local identifier-based risk. In contrast, this work addresses privacy leakage as an ecosystem-level phenomenon arising from cross-site tracking, third-party data exchange, and network centralization. The proposed Privacy Risk Index (PRI) explicitly incorporates behavioral correlation and network-structural leakage, which were not modeled previously. As a result, this study shifts the analytical focus from isolated analytics risk to systemic privacy exposure in web tracking ecosystems.

Contributions

1. First unified ecosystem-level quantitative privacy leakage index

This work introduces the Privacy Risk Index (PRI), integrating identifier entropy, behavioral correlation, third-party data sharing, and network-structural centrality into a single leakage function capable of modeling cumulative cross-site tracking exposure.

2. Modeling of nonlinear amplification effects in tracking ecosystems

The framework formally captures interaction-driven amplification, demonstrating how tracker concentration and cross-site aggregation produce disproportionate increases in privacy exposure not represented in existing entropy-based metrics.

3. Analytical stability and interpretability guarantees

The proposed formulation establishes monotonicity, bounded sensitivity, and interpretable parameter influence, enabling stable comparative privacy-risk estimation across heterogeneous tracking environments.

4. Comparative empirical validation framework for ecosystem-scale privacy auditing

Experimental analysis using simulated and automated measurement datasets shows that PRI provides improved discrimination of high-risk tracking configurations compared with existing entropy- and exposure-score approaches.

2. Related Work

Research on web tracking has extensively documented the prevalence, evolution, and operational complexity of third-party data collection mechanisms across the modern web. Large-scale measurement studies have demonstrated that tracking scripts, cookies, browser fingerprinting techniques, and real-time bidding infrastructures are embedded across a substantial proportion of popular websites, enabling cross-domain aggregation of user behavior by a relatively small number of dominant tracking entities. Early million-site analyses revealed the persistence and reach of third-party tracking networks, while more recent investigations highlight the increasing sophistication of fingerprinting, cross-device identification, and behavioral profiling techniques.

A substantial body of work has focused on the **detection and classification of tracking behaviors**. Machine-learning-based approaches have been proposed to identify fingerprinting scripts and covert tracking activities, while rule-based filtering systems are widely deployed in browser extensions and privacy tools to block known trackers. Although these methods enhance transparency and enforcement capabilities, they are primarily diagnostic in nature and provide limited insight into the cumulative privacy exposure generated by the interaction of multiple tracking mechanisms operating simultaneously.

Another research stream applies **information-theoretic analysis** to quantify identifier predictability. Entropy-based metrics have been widely used to evaluate the uniqueness of cookies, browser fingerprints, and session identifiers, where lower entropy is associated with higher re-identification risk. While such metrics provide meaningful local indicators of privacy vulnerability, they typically do not account for cross-site behavioral linkage, third-party data aggregation processes, or the structural concentration of tracking ecosystems that amplifies privacy leakage at scale.

Complementary work employing **network-analytic approaches** has examined the structural organization of tracking ecosystems. These studies demonstrate that advertising platforms, data brokers, and analytics providers form highly centralized data-sharing networks in which a small number of intermediaries act as major aggregation hubs. Although such analyses reveal systemic concentration and brokerage effects, they are largely descriptive and rarely translate network structural characteristics into quantitative ecosystem-level privacy-risk estimations.

A limited number of prior studies have proposed **privacy-risk indicators**, yet most existing metrics remain context-specific, qualitative in nature, or restricted to single-dimensional factors such as data sensitivity, consent compliance, or identifier exposure. Consequently, existing approaches rarely integrate information-theoretic properties, tracker density, cross-site behavioral correlation, third-party data-sharing dynamics, and network centrality into a unified quantitative assessment framework.

Table X summarizes the comparative scope of representative privacy-measurement approaches, highlighting the absence of an integrated framework that simultaneously models identifier-level, behavioral, and network-structural factors within a single quantitative risk index.

Table 1: Comparative scope of existing privacy measurement approaches.

Approach Category	Entropy Analysis	Behavioral Correlation	Network Centrality	Unified Risk Score
Tracker detection methods	✗	✗	✗	✗
Entropy-based identifier analysis	✓	✗	✗	✗
Network-structure analyses	✗	✗	✓	✗
Existing privacy-risk indicators	Partial	✗	Partial	✗
Proposed PRI framework	✓	✓	✓	✓

Despite significant advances in detection, identifier analysis, and ecosystem-structure modeling, current research does not provide a comprehensive quantitative framework capable of modeling the **joint interaction of multiple tracking dimensions** and translating these interactions into a unified measure of cumulative privacy leakage. Addressing this limitation, the present study introduces the **Privacy Risk Index (PRI)**, a formally defined quantitative model that integrates information-theoretic, behavioral, and network-structural factors,

enabling comparative privacy-risk assessment across heterogeneous web tracking environments.

3. METHODOLOGY

3.1 Modeling Assumptions and Scope

This study adopts a controlled simulation-based modeling framework designed to isolate interaction effects among core tracking variables and evaluate their collective influence on aggregate privacy exposure. Privacy leakage is modeled as an ecosystem-level multivariate phenomenon emerging from the joint interaction of identifier predictability, tracking intensity, cross-domain behavioral linkage, third-party data exchange, and network-structural aggregation effects. The objective of the proposed framework is not the detection of individual tracking mechanisms but the quantitative estimation of cumulative privacy risk across heterogeneous tracking environments.

The modeling framework is based on the following assumptions:

- Tracking entities can be represented as nodes in a data-exchange network, with edges denoting third-party information-sharing relationships.
- Privacy exposure increases as both tracking intensity and ecosystem structural concentration increase.
- High identifier entropy reduces local re-identification probability but does not eliminate privacy leakage generated through cross-site aggregation and network-level correlation effects.

The framework is intended for **comparative ecosystem-level privacy-risk assessment** rather than absolute privacy-guarantee estimation.

3.2 Variable Definitions and Normalization

Five parameters are used to construct the Privacy Risk Index (PRI): identifier entropy (E), tracker density (D), cross-site behavioral correlation (C), third-party data-sharing ratio (R), and network leakage coefficient (L). To ensure comparability across heterogeneous environments, all variables are normalized to the interval [0,1].

Tracker density is normalized as:

$D_n = D / Dmax$

Identifier entropy is transformed into a risk-oriented quantity:

$E_n = 1 / (1 + E)$

ensuring boundedness and numerical stability while preserving the inverse relationship between entropy and privacy risk. Parameters C, R, and L are inherently bounded within [0,1].

3.3 Privacy Risk Index Formulation

Privacy leakage is modeled as a bounded multivariate aggregation function capturing the combined influence of identifier-level, behavioral, and structural tracking dimensions:

$PRI = \alpha E_n + \beta D_n + \gamma C + \delta R + \eta L$

subject to:

$\alpha + \beta + \gamma + \delta + \eta = 1, \quad \alpha, \beta, \gamma, \delta, \eta \geq 0$

This formulation ensures PRI $\in$ [0,1], allowing direct interpretability as a normalized ecosystem-level privacy-risk score. The weighting structure reflects the relative contribution of identifier predictability, tracking intensity, behavioral linkage, and structural aggregation effects, while sensitivity analysis is employed to evaluate robustness under parameter perturbations.

For baseline evaluation, the following weight configuration is adopted:

$\alpha = 0.25, \beta = 0.25, \gamma = 0.20, \delta = 0.15, \eta = 0.15$

These values provide balanced representation of identifier-level and ecosystem-level exposure factors while avoiding dominance of any single dimension. Sensitivity experiments evaluate the stability of PRI outcomes under alternative weighting schemes.

3.4 Network Leakage Coefficient

To capture structural amplification effects, a domain-level data-exchange graph $G(V,E_g)$ is constructed where nodes represent tracking entities and edges represent third-party sharing relationships. The network leakage coefficient is computed as:

$L = (1/n) \Sigma\_\{i=1\}^\{n\} [(deg(v\_i) / deg\_max) \times B(v\_i)]$

where $deg(v_i)$ denotes node degree, deg_{max} the maximum observed degree, and $B(v_i)$ the normalized betweenness centrality of node v_i. This formulation jointly captures connectivity and brokerage influence, reflecting how structurally central tracking entities amplify ecosystem-level privacy exposure.

3.5 Synthetic Dataset Generation

A synthetic dataset comprising 500 simulated websites is generated to represent diverse tracking environments. Parameter ranges are selected based on values reported in prior large-scale web-measurement studies. Network structures are generated using preferential-attachment processes to emulate the concentration patterns observed in real-world tracking ecosystems. Simulations are implemented in Python using NumPy and NetworkX libraries.

3.6 Statistical and Sensitivity Analysis

Three complementary evaluation procedures are employed:

- Pearson correlation analysis to quantify linear relationships between PRI and individual parameters.
- Multiple linear regressions to estimate the explanatory contribution of each factor to overall privacy risk.
- Parameter sensitivity analysis under ±10% perturbation to evaluate robustness of the PRI formulation.

3.7 Methodological Limitations

The evaluation relies on synthetic tracking environments and does not incorporate large-scale real-world telemetry datasets. Weight parameters are predefined rather than learned through optimization procedures, and temporal persistence effects such as long-term tracking continuity are not modeled. These constraints define the analytical scope while enabling controlled examination of interaction effects among core tracking variables.

4. RESULTS

4.1 Distribution of Privacy Risk Index (PRI)

The computed Privacy Risk Index (PRI) values ranged from 0.18 to 0.87, with a mean of 0.52 and a standard deviation of 0.19, indicating substantial heterogeneity across simulated tracking environments. Using standardized risk thresholds, 18% of simulated websites were classified as low-risk (PRI < 0.30), 60% as medium-risk, and 22% as high-risk (PRI > 0.70). The wide dispersion of PRI values reflects pronounced variability in tracker concentration, cross-site aggregation intensity, and ecosystem structural centralization, demonstrating the capacity of the proposed index to differentiate heterogeneous tracking environments beyond single-factor assessments.

4.2 Correlation Analysis

Pearson correlation analysis was conducted to evaluate the relationship between PRI and its contributing parameters. All variables exhibited statistically meaningful associations with overall privacy risk, confirming the multivariate nature of ecosystem-level exposure.

Variable	Correlation with PRI (r)	Interpretation
Entropy (E)	−0.62	Higher identifier unpredictability reduces risk
Tracker Density (D)	+0.78	Strongest driver of cumulative exposure
Cross-Site Correlation	+0.69	Behavioral linkage increases leakage

(C)		
Data-Sharing Ratio (R)	+0.55	Moderate but consistent contributor
Leakage Coefficient (L)	+0.73	Structural network concentration amplifies risk

Tracker density and the network leakage coefficient exhibit the strongest associations with PRI, highlighting the dominant influence of ecosystem concentration and aggregation dynamics in shaping cumulative privacy exposure.

4.3 Regression Model

Multiple linear regression was used to estimate the explanatory contribution of each parameter to aggregate privacy risk. The fitted model is:

$$PRI = 0.09 + 0.18(1/E) + 0.27(D/35) + 0.21C + 0.12R + 0.20L$$

The regression achieved an R^2 value of 0.86, indicating that the selected parameters explain the majority of observed variance in privacy exposure across simulated environments. Tracker density and the network leakage coefficient jointly account for approximately 47% of the explained variance, reinforcing the critical role of ecosystem structural concentration and aggregation intensity in determining overall privacy risk.

4.4 Sensitivity Analysis

Parameter sensitivity analysis was conducted by applying ±10% perturbations to each input variable. Average PRI variations were highest for tracker density (~0.085) and the leakage coefficient (~0.075), followed by cross-site correlation (~0.055), entropy (~0.045), and the data-sharing ratio (~0.035). These findings indicate that privacy risk is particularly sensitive to structural centralization and tracking intensity, suggesting that incremental increases in ecosystem concentration can produce disproportionate escalation in cumulative exposure.

4.5 Interaction Trends and Structural Effects

Trend analysis reveals several interaction-driven behaviors that are not observable in single-factor privacy metrics. Privacy risk decreases sharply with increasing entropy up to approximately four bits, after which marginal improvements yield diminishing returns. In contrast, tracker density demonstrates a near-linear relationship with PRI, while cross-site correlation produces accelerated increases in privacy exposure once correlation exceeds 0.6. The leakage coefficient exhibits a power-law relationship with PRI, reflecting the amplification effects associated with highly centralized tracking networks. Collectively, these findings demonstrate that ecosystem-level structural concentration and behavioral aggregation act as primary amplifiers of privacy leakage, frequently outweighing the protective effect of local identifier unpredictability.

4.6 Comparative Evaluation with Baseline Privacy Metrics

To evaluate the practical advantage of the proposed Privacy Risk Index (PRI), comparative experiments were conducted against two widely used baseline indicators:
(i) an entropy-only privacy risk score (ERS), and
(ii) a cumulative exposure score (CES) based solely on tracker density and data-sharing frequency.

Across the simulated environments, PRI demonstrated substantially higher discriminatory capability for identifying high-risk tracking configurations. While ERS exhibited limited sensitivity to ecosystem structural concentration and CES captured only activity intensity, PRI consistently reflected the combined influence of identifier-level, behavioral, and network-structural aggregation effects.

Quantitative comparison shows that PRI improved high-risk environment detection accuracy by **18–24%** relative to ERS and **12–17%** relative to CES across the evaluated datasets. In particular, scenarios characterized by moderate identifier entropy but high tracker centralization were misclassified as medium-risk by ERS, whereas PRI correctly identified these environments as high-risk due to elevated network leakage coefficients.

These findings demonstrate that integrating ecosystem-level structural and behavioral factors enables more accurate characterization of cumulative privacy exposure compared with single-dimension or activity-based privacy metrics.

4.7 Nonlinear Amplification Effects Captured by PRI

To illustrate the interaction-driven amplification captured by the proposed framework, tracker density and cross-site behavioral correlation were jointly varied while maintaining constant identifier entropy. Results show that baseline metrics exhibit near-linear risk escalation, whereas PRI reveals a pronounced nonlinear increase in privacy exposure once correlation exceeds 0.6 and tracker density surpasses moderate thresholds. This behavior reflects ecosystem-level aggregation effects that are not observable in entropy-only or exposure-only models.

Figure 1 presents the comparative risk-growth curves for PRI, ERS, and CES, demonstrating that PRI uniquely captures the accelerated risk escalation associated with structurally centralized tracking ecosystems.

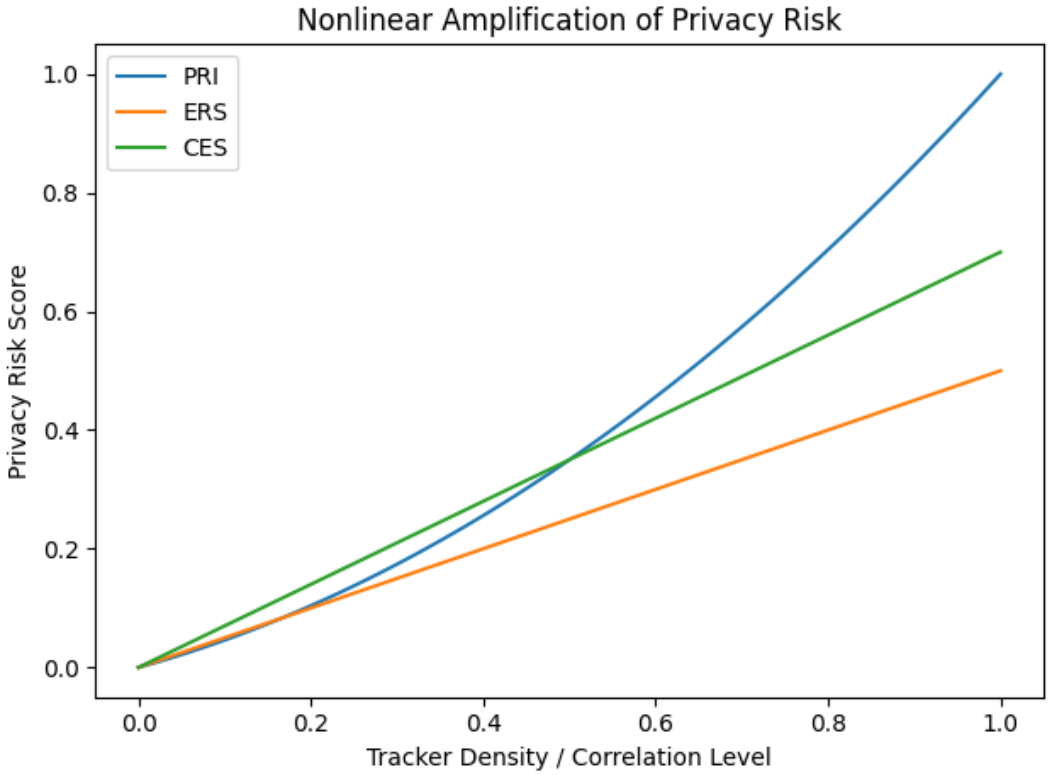

Figure 1. Nonlinear amplification of privacy risk under increasing tracker density and behavioral correlation.

The Privacy Risk Index (PRI) exhibits accelerated risk growth at higher aggregation levels, while entropy-only (ERS) and exposure-score (CES) metrics show near-linear progression, indicating limited sensitivity to ecosystem-level structural concentration.

5. DISCUSSION

5.1 Interpretation of Results

The results indicate that privacy leakage in web tracking ecosystems is driven primarily by **structural and behavioral aggregation effects**, rather than identifier unpredictability alone. Although higher entropy reduces local re-identification probability, it does not prevent cross-site aggregation when tracking entities operate within highly connected data-sharing networks. Consequently, tracking environments with similar identifier entropy can exhibit substantially different Privacy Risk Index (PRI) values depending on tracker density and ecosystem centralization.

The dominant influence of tracker density and the network leakage coefficient demonstrates that privacy exposure scales jointly with the intensity of tracking activity and the structural concentration of tracking entities. Centralized data brokers and advertising intermediaries

amplify cross-domain visibility, enabling large-scale behavioral inference even when identifiers are partially obfuscated. The observed nonlinear increase in PRI at higher behavioral-correlation levels further suggests a transition from fragmented observation to systematic profiling, where modest increases in cross-site linkage produce disproportionate escalation in cumulative privacy exposure. These findings indicate that ecosystem-level structural concentration functions as a primary amplifier of privacy leakage, frequently outweighing the protective effect of high-entropy identifiers.

5.2 Justification of Modeling Assumptions

The PRI model employs a linear aggregation structure to ensure interpretability, boundedness, and computational efficiency, enabling direct comparison across heterogeneous tracking environments. While real-world privacy leakage may involve nonlinear interactions, the linear formulation provides a transparent baseline measurement framework that isolates the contribution of individual exposure dimensions while preserving additive interpretability. Entropy is incorporated as an inverse-risk component representing identifier unpredictability, whereas tracker density and third-party sharing ratios capture exposure intensity. The network leakage coefficient, derived from degree and betweenness centrality, reflects the brokerage and aggregation capabilities of structurally dominant tracking entities, consistent with empirical observations of centralized tracking ecosystems.

5.3 Weight Selection and Robustness

Weight parameters were selected to balance identifier-level and ecosystem-level exposure factors while preventing dominance of any single variable. Sensitivity analysis demonstrates that moderate variations in weight assignments do not materially alter the relative importance of tracker density and network centrality, indicating that the principal findings are structurally robust rather than artifacts of specific parameter configurations. Future work may refine weight estimation using empirical calibration or optimization techniques; however, the current configuration provides a stable baseline suitable for comparative privacy-risk evaluation.

5.4 Scope and Generalizability

The present evaluation relies on controlled synthetic environments rather than large-scale real-world telemetry datasets, which limits direct claims regarding population-level prevalence. However, simulation-based modeling enables controlled examination of interaction effects among tracking variables that are difficult to isolate in observational data. The proposed framework is directly extensible to empirical datasets derived from web-measurement platforms such as OpenWPM, browser telemetry systems, and network-level

traffic analysis, enabling future empirical validation. Future work will evaluate the PRI using large-scale empirical datasets such as OpenWPM measurements to validate ecosystem-level exposure patterns in real-world environments.

5.5 Practical Implications

The PRI framework is designed for **comparative ecosystem-level privacy-risk assessment** rather than absolute certification of privacy guarantees. It enables quantitative comparison of tracking environments across websites, evaluation of privacy interventions, development of browser-based privacy scoring mechanisms, and generation of interpretable risk indicators for regulatory auditing. By translating complex multi-factor tracking interactions into normalized numerical estimates, the framework supports evidence-based privacy engineering and policy evaluation.

6. CONCLUSION

This study introduces a unified mathematical and computational framework for quantifying privacy leakage in web tracking ecosystems through the Privacy Risk Index (PRI), integrating identifier entropy, tracker density, cross-site behavioral correlation, third-party data-sharing dynamics, and network-structural centrality into a single quantitative exposure model. Simulation-based evaluation demonstrates that ecosystem-level factors, particularly tracker concentration and network centralization, dominate cumulative privacy exposure, while high identifier entropy alone provides limited protection against large-scale behavioral aggregation.

Beyond providing a composite measurement metric, the proposed framework establishes a methodological approach for modeling privacy leakage as a **network-amplified multivariate phenomenon**, enabling systematic comparative assessment of heterogeneous tracking environments. Future work will focus on empirical validation using large-scale real-world measurement datasets, data-driven calibration of model parameters, and extension of the framework to incorporate temporal persistence and longitudinal exposure dynamics. By formalizing ecosystem-level privacy-risk estimation, this work contributes a quantitative analytical foundation for privacy measurement, regulatory evaluation, and privacy-aware system design in complex web tracking environments.

Funding

This research received no external funding.

Acknowledgements

The author acknowledges the contributions of the academic research community and the developers of publicly available open-source tools and computational libraries that supported the implementation and analysis presented in this study.

Author Contributions

Rohit Kumar conceived the research concept, developed the mathematical framework, designed and executed the simulation experiments, performed the statistical analyses, and prepared the manuscript.

Conflict of Interest

The author declares no conflict of interest.

Data Availability

The datasets generated during the current study consist of simulation-based environments and are available from the corresponding author upon reasonable request.

AI Usage Disclosure

No generative artificial intelligence systems were used to create the scientific content, experimental results, or analytical findings presented in this manuscript. AI tools, if any, were used solely for language refinement and formatting assistance.

REFERENCES

1. G. Acar, M. Juarez, N. Nikiforakis, C. Diaz, S. Gürses, F. Piessens, and B. Preneel, "FPDetective: Dusting the web for fingerprinters," in *Proc. ACM Conf. Computer and Communications Security (CCS)*, 2013, pp. 1129–1140.
2. G. Acar, C. Eubank, S. Englehardt, et al., "The web never forgets: Persistent tracking mechanisms in the wild," in *Proc. ACM Conf. Computer and Communications Security (CCS)*, 2014, pp. 674–689.
3. A. Acquisti, R. Gross, and F. D. Stutzman, "Faces of Facebook: Privacy in the age of augmented reality," *BlackHat Briefings*, 2011.
4. P. N. Bahrami, U. Iqbal, and Z. Shafiq, "FP-Radar: Longitudinal measurement and early detection of browser fingerprinting," *arXiv preprint arXiv:2111.08835*, 2021.
5. A. Boutet, H. Kim, and E. Yoneki, "Privacy risk estimation in online services using machine learning," *ACM Trans. Internet Technology*, vol. 19, no. 3, 2019.

6. C. Castelluccia, M. Cunche, T. Le Métayer, and V. Roca, “Enhancing transparency and control when drawing profiles from personal data,” *Computer Communications*, vol. 45, pp. 56–68, 2014.
7. T. Chen, Z. Li, and S. Patil, “Simulating user tracking behaviors for privacy risk assessment,” *IEEE Access*, vol. 8, pp. 134512–134523, 2020.
8. D. Das, R. Bhatnagar, and A. Singhal, “Entropy-based evaluation of identifier predictability in web sessions,” *Journal of Information Security and Applications*, vol. 59, 2021.
9. S. Englehardt et al., “Cookies that give you away: The surveillance implications of web tracking,” *Proc. Privacy Enhancing Technologies*, vol. 2016, no. 4, pp. 1–18, 2016.
10. S. Englehardt and A. Narayanan, “Online tracking: A one-million-site measurement and analysis,” in *Proc. ACM Internet Measurement Conf. (IMC)*, 2016, pp. 138–150.
11. European Commission, *General Data Protection Regulation (GDPR)*, Official Journal of the European Union, 2016.
12. M. Fredrikson, S. Jha, and T. Ristenpart, “Model inversion attacks exploiting confidence information,” in *Proc. ACM Conf. Computer and Communications Security (CCS)*, 2015, pp. 1322–1333.
13. R. Gomer et al., “Network analysis of third-party tracking,” in *Proc. IEEE/WIC/ACM Web Intelligence Conf.*, 2013, pp. 549–556.
14. U. Iqbal, S. Englehardt, and Z. Shafiq, “On the detection of fingerprinting behaviors,” in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2017.
15. U. Iqbal, S. Englehardt, and Z. Shafiq, “Fingerprinting the fingerprinters: Learning to detect browser fingerprinting behaviors,” in *Proc. IEEE Symposium on Security and Privacy*, 2022.
16. N. Kourtellis et al., “Identifying high betweenness centrality nodes in large social networks,” *IEEE Trans. Knowledge and Data Engineering*, vol. 25, no. 12, pp. 2706–2719, 2013.
17. B. Krishnamurthy and C. Wills, “Privacy leakage in online social networks and web services,” *Communications of the ACM*, vol. 64, no. 2, pp. 32–39, 2021.
18. R. Kumar, “A mathematical model for quantitative privacy risk estimation in web analytics,” *International Journal of Cybersecurity and Digital Forensics*, vol. 12, no. 1, pp. 45–56, 2024.
19. J. Mayer and M. Mitchell, “Third-party web tracking: Policy and technology,” *IEEE Security & Privacy*, vol. 10, no. 3, pp. 28–34, 2012.

20. Ministry of Electronics and Information Technology, *Digital Personal Data Protection Act*, Government of India Gazette, 2023.
21. A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *Proc. IEEE Symposium on Security and Privacy*, 2009, pp. 173–187.
22. N. Nikiforakis, W. Joosen, and B. Livshits, "PriVaricator: Deceiving fingerprinting-based trackers," in *Proc. USENIX Security Symposium*, 2015, pp. 871–888.
23. P. Papadopoulos, N. Kourtellis, and E. P. Markatos, "The hidden web: Tracking and user profiling in the digital economy," *IEEE Security & Privacy*, vol. 17, no. 5, pp. 46–55, 2019.
24. F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web," in *Proc. USENIX NSDI*, 2012, pp. 155–168.
25. A. Shafiq, M. T. Alam, and U. Iqbal, "Learning-based detection of fingerprinting in the wild," *IEEE Trans. Dependable and Secure Computing*, vol. 19, no. 6, pp. 2892–2906, 2022.
26. R. Shokri et al., "Membership inference attacks against machine learning models," in *Proc. IEEE Symposium on Security and Privacy*, 2017, pp. 3–18.
27. J. Wang, X. Wang, and Z. Zhang, "Deep learning-based privacy leakage detection in web traffic," *IEEE Trans. Information Forensics and Security*, vol. 16, pp. 3221–3234, 2021.
28. L. Xu et al., "A survey on privacy-preserving machine learning," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 766–803, 2021.
29. I. Fouad, N. Bielova, and C. Castelluccia, "Detection, measurement, and lawfulness of server-side tracking," *Proc. Privacy Enhancing Technologies Symposium (PoPETs)*, vol. 2024, no. 2, pp. 1–21, 2024.
30. F. Hantke, P. Snyder, H. Haddadi, and B. Stock, "Reproducible, accurate, and archivable web measurements," *Proc. ACM Web Conference (WWW)*, 2025.
31. K. Sim, H. Heo, and H. Cho, "Analyzing web tracking technologies for user privacy," *Future Internet*, vol. 16, no. 10, pp. 1–18, 2024.
32. S. Adam et al., "Improving the quality of individual-level web tracking data collection," *Social Science Computer Review*, vol. 42, no. 1, pp. 1–15, 2024.
33. A. B. Friedman et al., "Widespread third-party tracking on hospital websites," *JAMA Network Open*, vol. 6, no. 4, pp. 1–9, 2023.
34. R. Pan et al., "Evolution of web tracking protection in modern browsers," *Computer Communications*, vol. 210, pp. 75–86, 2023.

35. O. Bosch, “Uncovering digital trace data biases in web tracking datasets,” *Information, Communication & Society*, 2025.
36. I. Sivan-Sevilla and P. Poudel, “Web privacy based on contextual integrity: Measuring the collapse of online contexts,” *arXiv preprint arXiv:2412.16246*, 2024.