I Know What You Bought Last Summer: Investigating User Data Leakage in E-Commerce Platforms

Ioannis Vlachogiannakis[1,2], Emmanouil Papadogiannakis[1,2], Panagiotis Papadopoulos[1], Nicolas Kourtellis[3] and Evangelos Markatos[1,2]

[1]*Foundation for Research and Technology - Hellas (FORTH), Heraklion, Greece*

[2]*University of Crete, Heraklion, Greece*

[3]*Telefonica Research, Barcelona, Spain*

Keywords: Private Information Leakage, Sensitive User Data, E-Commerce, Privacy, Tracking.

Abstract: In the digital age, e-commerce has transformed the way consumers shop, offering convenience and accessibility. Nevertheless, concerns about the privacy and security of personal information shared on these platforms have risen. In this work, we investigate user privacy violations, noting the risks of data leakage to third-party entities. Utilizing a semi-automated data collection approach, we examine a selection of popular online e-shops, revealing that nearly 30% of them violate user privacy by disclosing personal information to third parties. We unveil how minimal user interaction across multiple e-commerce websites can result in a comprehensive privacy breach. We observe significant data-sharing patterns with platforms like Facebook, which use personal information to build user profiles and link them to social media accounts.

1 INTRODUCTION

According to studies (Forbes, 2024), 34% of shoppers shop online at least once a week. In general, the e-commerce market is expected to reach a staggering $8 trillion value by 2027. Apart from the apparent comfort of shopping without visiting physical stores, one of the key factors of the rapid growth of e-commerce has been the extensive use of data and third-parties (i.e., analytics, media buttons, advertising networks).

By integrating such third-party services, e-shops can optimize inventory levels by analyzing historical sales data, identifying seasonal trends, and forecasting future demand. Additionally, and more importantly, by collecting and processing a great wealth of user and behavioral data, third-party analytics can provide a unique understanding of customer behavior and their purchase patterns (Group, 2024). This allows e-shops to increase customer retention and perform targeted advertising (Fabbro, 2024).

In some cases, e-shops, may not have full control or even awareness of what/how pervasive the tracking of the third-party tools they embed in their platforms is. On the other hand, customers can only trust their sensitive personal information (e.g., contact details and payment information) to e-shops and expect that this information will be used only by them and for the sole purpose of purchasing products or services.

Unfortunately, there are various examples where this trust was breached. In 2025, an Austrian privacy non-profit filed complaints accusing e-shops like AliExpress, SHEIN, Temu for violating data protection regulations in the European Union by unlawfully transferring users' data to China (Lakshmanan, 2025). In 2024 the state of Arkansas sued the Chinese online retailer Temu for illegally accessing user information (Smith, 2024). In 2023, hundreds of online stores were reported for accidentally leaking customer data in public folders without any restrictions (Weigand, 2023). In 2019, a study showed that at least 80% of shopping apps leak users' data (Langone, 2019).

Prior research on e-commerce has highlighted several vulnerabilities in online shopping platforms, including weak API security (Flores et al., 2022), third-party tracking (Rauti et al., 2024), and insufficient data protection measures (Pagey et al., 2023). Previous works have identified that e-commerce platforms share user data, but do not provide a comprehensive mapping of how information flows between different third-party entities. Data leaks are often examined in isolation (i.e., individual e-shop platforms), ignoring the aggregation of user data across multiple platforms. A more comprehensive approach is essential to map the full lifecycle of leaked data globally.

In this work, we investigate the leakage of sensitive user information from e-commerce platforms to

Vlachogiannakis, I., Papadogiannakis, E., Papadopoulos, P., Kourtellis, N. and Markatos, E.
I Know What You Bought Last Summer: Investigating User Data Leakage in E-Commerce Platforms.
DOI: 10.5220/0013519000003979
In *Proceedings of the 22nd International Conference on Security and Cryptography (SECRYPT 2025)*, pages 392-399
ISBN: 978-989-758-760-3; ISSN: 2184-7711

third parties and explore how these entities can aggregate user data across multiple e-shops. We explore whether the personal information that users provide to e-commerce platforms are shared with third parties, contrarily to the user's expectations. Our findings highlight that privacy leaks are not limited to obscure sites, but extend to highly popular e-shops with million of monthly visitors. Additionally, even limited interaction with multiple platforms can lead to complete exposure of a user's personal information.

The main contributions of this work are:

1. We compare the data-leaking behaviors of e-shops with those in other industries, emphasizing the heightened privacy risks in the e-commerce and shopping field.
2. We discover that 29% of the online retail stores in our dataset share at least one piece of their users' sensitive private information to a third-party entity. We highlight that this behavior is evident even in extremely popular platforms with millions of monthly visitors.
3. We demonstrate that third parties are capable of aggregating personal information from multiple e-commerce platforms to construct comprehensive user profiles. In fact, users engaging with as few as five e-commerce sites may have their entire profile exposed to third parties.
4. We reveal that Meta is the third party that receives the largest amount of private information, allowing it to use this information to match shopping behaviors with Facebook accounts.
5. We make our tools and dataset publicly available to foster further research on the field.

2 METHODOLOGY

In this work, we follow a two-phase methodology. The first phase involves data collection, where we emulate real-world scenarios to gather data from e-commerce platforms. The second phase involves data analysis, where we process the collected information. For data collection, we develop a semi-automated crawler using the Playwright framework (Microsoft, 2020), to systematically extract data from the websites of selected e-shop platforms.

2.1 Data Collection

First, we compile a list of popular and representative e-commerce websites from around the world. To that extent, we utilize SimilarWeb (Similarweb-LTD, 2025) to gather the most popular five e-shop platforms in each country along with the top 50 worldwide, and accumulate a list of 200 distinct e-commerce websites. We make our list publicly available to foster further research on the field (Vlachogiannakis, 2025). Then, we visit all the e-commerce platforms with our semi-automated crawler, located in an EU institute.

We build our tool to manage a browser instance in a way that it collects all network traffic and cookie jar (both first-party and third-party cookies), as the user interacts with the browser. Specifically, for e-shop websites, these actions involve creating a user account, browsing products, and preparing for a purchase. Our scenario simulates real user activities such as product search, adding items to the cart, and progressing through steps leading up to payment and order confirmation. We deliberately refrain from completing any purchases to avoid impacting the platforms or merchants in any way. Ethical considerations are further discussed in Section 6.3.

Our goal is to emulate a real-world case reflecting a typical user with unique personal information such as full name, mobile phone, email address, physical address, etc. Towards that extent, we create a fake persona of a user from the country where the crawler is located. It is important to highlight that the persona is consistent and in each platform we provide the same personal information in the same manner as a real user. Our persona consists of the following personal and sensitive information: (1) email, (2) name, (3) phone number, (4) gender, (5) zip code, (6) credit card details, (7) username, and (8) password.

2.2 Leakage Detection

We manually visit all the websites in our list during July 2024 independently of each other, starting with a clean browser context for each e-commerce platform. We extensively study the network traffic generated during our visit to each e-commerce platform. To detect e-commerce websites that leak sensitive information, we study how information flows from websites to third parties. We group HTTP(S) requests and filter them by destination URLs to identify all third parties, with which the platforms communicate. Additionally, we make use of the DuckDuckGo Tracker Radar dataset (DuckDuckGo, 2020) to match each one of the third-party domains with the company that owns or operates them, aiming to link information leaks with specific companies or larger legal entities.

Finally, we iterate through all HTTP(S) requests and third-party cookies, searching for occurrences of all sensitive personal information that we inserted when creating each profile. We search for this in-

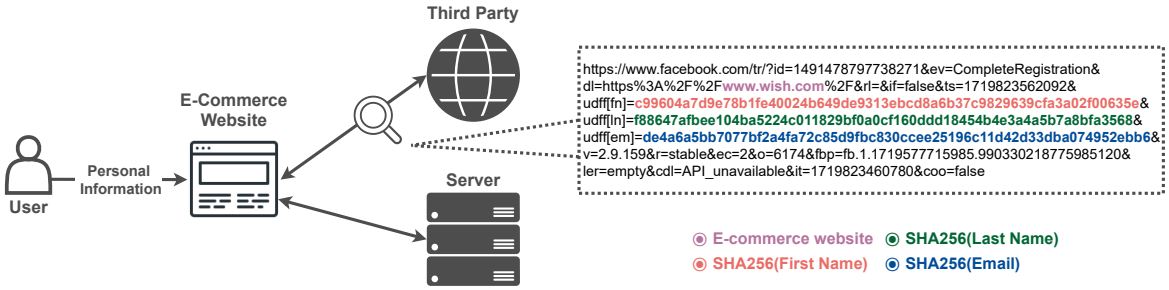

Figure 1: Overview of methodology for detecting personal information leakage.

formation in different formats, including plaintext, SHA256 or MD5 Hashed, URL Encoded and Base64 Encoded. We provide an overview of our methodology in Figure 1, where we demonstrate how we capture personal information being shared with third parties. We conduct an in-depth analysis of the collected data to unveil matches in the URL parameters of GET requests and the body of POST requests, as well as in each cookie value. We study information flows between the client browser and third parties, acknowledging that data sent directly from the store's server to third parties is not captured by our methodology.

Website operators might claim that sharing hashed personal data with third parties (as shown in Figure 1) is harmless, since the original data cannot be extracted. However, we argue that this is a misleading argument. Third parties, such as analytics services and social networks, with access to vast amounts of user data, can easily link hashed values to their own databases. If a service already holds a user's email address, it can effortlessly identify the individual by matching the hash to its database entry.

3 SENSITIVE INFORMATION LEAKAGE

Inspired by anecdotal evidence that e-commerce platforms collect an extensive amount of user information (Smith, 2024; Group, 2024; Fabbro, 2024), we perform a preliminary investigation to understand to what extent this is happening. Our goal is to discover indications that private information leaks are more common in e-commerce platforms since they have access to more user information compared to other types of websites. We develop a data collection tool designed to automatically visit websites and capture network traffic along with the cookie jar, including both first-party and third-party cookies. The tool navigates to a website's landing page and waits for the page to fully load before collecting the relevant data. We select five different categories of websites to study, (i) "E-commerce and Shopping", (ii) "Business and Consumer services", (iii) "Health", (iv) "Travel and Tourism", and (v) "Finance". For this experiment, we process 200 websites from each category extracted from SimilarWeb (Similarweb-LTD, 2025). We make the lists of websites per category publicly available (Vlachogiannakis, 2025). We collect all requests and cookies, extract third-party entities, and compare the number of interactions to determine which category engages with the highest number of third parties.

Table 1: Average, Median and 90th Percentile of third-party interactions. E-commerce websites interact with more third-party services than other categories.

Category of Websites	Average	Median	90th Percentile
E-commerce & Shopping	17	12	40
Business & Consumer	15	12	33
Health	15	12	34
Travel & Tourism	14	10	35
Finance	13	11	31

Our analysis reveals that websites in the "E-commerce and Shopping" category interact with more third-party services than those in other categories. We present our findings in Table 1, showing the average, median and 90th percentile of third-party interactions across five categories. A deeper analysis of e-commerce platforms from our original list (see Section 2.1) reveals that the average and median number of third parties can rise to 21 and 14, respectively, when users spend more time on the website, interacting with its components and navigating to more pages apart from the landing page. Previous work has already demonstrated that landing and internal pages can have significant differences in the number of trackers (Aqeel et al., 2020).

> **Finding 1:** E-commerce websites engage with more third-party services than other categories, with the number of interactions increasing as users spend more time navigating the site.

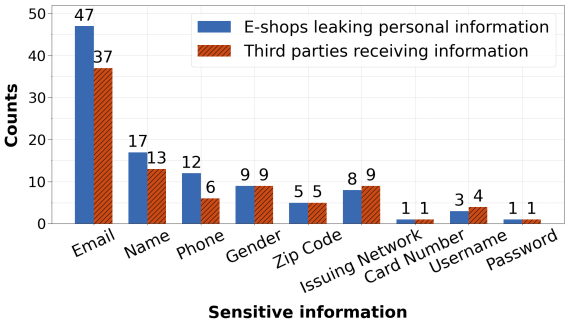

Figure 2: Number of e-shops leaking sensitive personal information and number of third parties collecting this information from different e-commerce platforms.

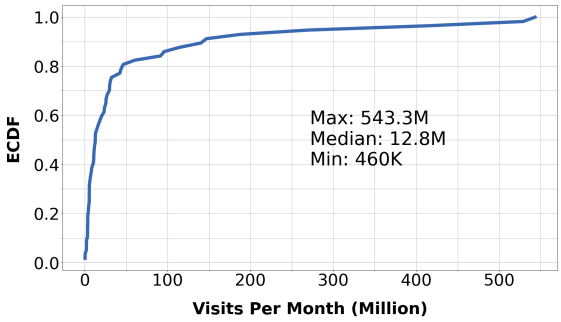

Figure 3: Distribution of monthly visits (both desktop and mobile) of e-commerce websites.

3.1 Sensitive Data Flows

We analyze the collected data following the methodology outlined in Section 2 to identify instances of sensitive information leakage from e-commerce platforms to third parties. We discover that 57 out of 200 digital shopping platforms we investigated, almost 30%, leak at least one piece of sensitive user information to an external legal entity. This means that almost one in three online stores transmits sensitive user data, either encoded or in plain text, to unrelated third parties. Unlike pseudonymous tracking methods such as third-party cookies or browser fingerprinting (Papadogiannakis et al., 2021), this type of data leakage is particularly concerning because it involves personally identifiable information (PII), including full names, email addresses, and physical addresses. The collection of such information from third parties not only compromises user privacy, by enabling detailed profiling, but also increases the risk of exposure in cases of data breaches, a common occurrence in the last few years (e.g., (Klappholz, 2024; Firstpost, 2024)).

In Figure 2, we illustrate the number of e-commerce platforms sharing personal information (blue bars), as well as the third party entities collecting user information from various e-commerce platforms (red bars). We observe that the email address, a piece of sensitive information that uniquely identifies a user, is leaked from 47 online stores. Also, it is worth noting that 37 distinct third parties collect this information and can, at the very least, correlate where and when a specific user (i.e., email address) shops online. These third parties include popular conglomerates that provide analytics services (e.g., Facebook, Google) as well as companies that can correlate shopping profiles with user accounts in other platforms (e.g., ByteDance Ltd that developed TikTok). Third-party data brokers buy and compile information from multiple sources, often without users' knowledge. If separate e-commerce platforms leak partial user information, data brokers can combine these fragments to create complete profiles, including names, addresses, purchase history, and even preferences.

Next, we study the popularity of the 57 retail online platforms leaking user information. Figure 3 illustrates their popularity distribution, which is based on number of monthly visits, both from mobile and desktop clients, obtained from SimilarWeb. Our analysis reveals that e-commerce platforms leaking user information range from low-visibility websites to those with substantial monthly traffic. While it is somewhat expected that less popular online stores may engage in such practices due to limited regulatory oversight, we also observe this behavior in highly popular platforms with millions of monthly visitors, including AliExpress and Etsy. Altogether, e-commerce platforms that share at least one piece of personal information have an aggregated traffic of *3.23* Billion monthly visits, significantly increasing the risk of user data exposure on a massive scale.

> **Finding 2:** Our analysis reveals that 29% of the online retail stores in our dataset, including highly popular platforms with millions of monthly visitors, leak at least one piece of their users' sensitive private information.

3.2 Data Aggregation

Numerous third-party tracking services, such as Google Analytics, Meta Pixel, and various advertising networks, aggregate data from multiple e-commerce platforms. When a user engages with sites like AliExpress and Wayfair, these tracking entities can correlate their activity across platforms, enabling the construction of comprehensive consumer profiles. We discover that a user visiting as few as five different e-shop platforms, can have their entire personal profile (including contact information, account credentials, payment details, etc.) shared with third parties, as illustrated in Figure 4. This personal information,

which includes sensitive fields such as email address, name, and phone number collectively forms a comprehensive user profile, as described in Section 2.1. This suggests that even limited interactions with e-shops can pose significant privacy risks, as users are unknowingly subjected to data sharing without their explicit consent.

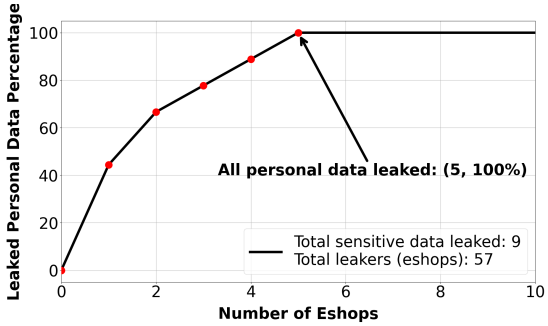

Figure 4: Complete exposure of a user's personal information when visiting as few as 5 e-shop platforms.

We aggregate popular third-party entities that receive personal information from the e-shops we study, and present them in Table 2, along with the information they collect. It is evident that Web conglomerates such as *Facebook*, *Google* and *Microsoft* not only collect very sensitive personal information, but they do so from multiple online stores, thus tracking users when they shop in various platforms. We illustrate the flow of sensitive user information towards third parties in Figure 5. This plot illustrates the most critical fields of personal information and the third parties that receive this data. Each flow, represented by its width, indicates the volume of e-shops leaking a piece of personal information to a third party. A wider flow suggests a higher number of e-commerce platforms sharing sensitive data. We observe that the email address, a unique identifier, is commonly leaked by e-shops. Moreover, it is evident that Facebook collects the most data from e-commerce platforms.

Table 2: Third-party legal entities acquiring the most personal information from multiple e-commerce platforms. Each cell represents the number of distinct e-shops sharing specific personal information with each third-party entity.

Third-Party Company	Email	Name	Phone	Gender
Facebook, Inc.	37	14	9	1
Google LLC	12	3	-	3
ByteDance Ltd. (TikTok)	12	2	5	1
Microsoft Corporation	3	3	-	1
Snap Inc. (Snapchat)	6	1	-	-

It is worth noting that when emulating the real world scenario described in Section 2, we browse the same category of products wherever possible (e.g., shoes) in each platform and add them to the virtual cart. We discover that some e-shop platforms inform third parties about the products a user is interested in. In Listing 1, we demonstrate a case of a decoded URL informing Facebook that the user browsed for a specific category of products.

Many e-commerce platforms share user data with third-party trackers, which operate across multiple websites. These trackers can link user activity from different platforms to create detailed consumer profiles, including sensitive information like names, addresses, and purchase history - often without users' awareness. This creates significant privacy risks, allowing third parties to build invasive profiles of their online activities and personal preferences

Finding 3: Users interacting with as few as five e-commerce sites risk having their entire profile exposed to third parties that consolidate personal information to create detailed user profiles.

4 EFFECTIVE PERSISTENT TRACKING

We observe in Figure 5 that there is a significant flow of sensitive information from retail platforms towards Facebook. Upon closer inspection, we find that 18.5% of the e-shop platforms we studied leak the user's email address to Facebook. This information is shared to Facebook as a hashed value, often along other fields like username. Facebook, one of the largest social networks, has an extensive collection of user email addresses. Through tracking on e-shops, Facebook can link users browsing e-shops to specific Facebook accounts. This form of tracking is particularly effective, as email addresses (unlike pseudonymous third-party cookies) are unique and directly identifiable.

To make matters worse, we observe that when e-shop platforms send requests towards Facebook's endpoints, the product or category of products that the user is browsing is also leaked (Listing 1). Facebook is, therefore, capable of tracking users' shopping behavior by identifying products they have seen or bought and build an extensive user profiles. As a result, Facebook not only *knows where and when you are shopping, but also what you are shopping for*.

We argue that this form of tracking has a less apparent dimension. Facebook is able to track users through its tracking services, that retail online stores integrate to their platforms, even without owning a Facebook account. Once the user decides to create an account, Facebook can associate all previously collected personal and historical data with the newly cre-

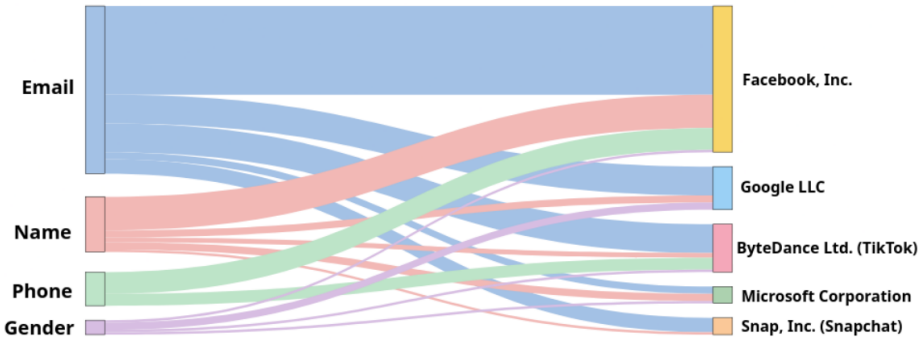

Figure 5: Information flow of sensitive personal information that e-commerce platforms distribute to third-party entities. A greater flow weight indicates that a third party receives information from multiple online stores.

```
https://www.facebook.com/tr/?id=693212574061933&ev=PageView&dl=https://www.
    thewarehouse.co.nz/c/clothing-shoes-accessories/mens-clothing-shoes/
    mens-shoes/mens-sports-hiking-shoes&rl=https://www.thewarehouse.co.nz/account&
    if=false&ts=1719495540857&sw=1920&sh=1200&ud[em]=de4a6a5bb7077bf2a4fa72c
85d9fbc830ccee25196c11d42d33dba074952ebb6&v=2.9.159&r=stable&a=tmSimo-GTM-
    WebTemplate&ec=0&o=4126&fbp=fb.2.1719495152216.771782349158911l8&ler=empty
    &it=1719495540626&coo=false&eid=1719495539212_1719496172506_1_gtm.js&tm=1
```

Listing 1: Destination URL captured in one of the requests in our dataset. Its destination is the tracking service of Facebook and as a parameter is passed the exact url of the product that our virtual user visited.

ated profile. This enables Facebook to gain insights into the *user's past shopping habits and preferences*. As a result, the company acquires a comprehensive understanding of consumer behavior, allowing it to offer personalized recommendations and advertisements. To put it into perspective, when a new user creates an entirely new account with Facebook, the company may already be aware of their shopping habits. This behavior has been similarly noted in previous research involving Facebook's Pixel tracking technology (Bekos et al., 2023).

Finally, Facebook is part of the Meta group, which also operates Instagram and WhatsApp, thus broadening the scope of data collection across multiple social platforms. This interconnected network grants Meta comprehensive insights into user preferences and shopping behaviors. By leveraging data from Facebook, WhatsApp, and Instagram, Meta can effectively track a diverse range of users, often segmented by age groups linked to each platform (Center, 2024). The cumulative traffic across the 37 e-shops, that disclose users' emails to Facebook, is *2.35* Billion, highlighting the substantial reach and potential privacy impact of these data sharing practices.

Finding 4: Meta is the third party receiving the most significant amount of private information, enabling the company to correlate shopping behaviors with specific Facebook accounts.

5 RELATED WORK

In (Okeke et al., 2013), the authors highlighted privacy and trust concerns among online customers regarding data security and sharing personal information with third parties. In (Gurung and Raja, 2016), the authors suggest that privacy concerns have a greater impact on risk assessment than security concerns, influencing consumer attitudes and intentions toward online shopping. In (Broeder, 2020), the authors found that privacy notices indirectly influenced trust and purchase behavior by assuring consumers of personal information protection. Gaining their satisfaction and trust leads customers to prioritize online shopping against traditional shopping methods, as authors in (Kurniawan and Setyawan, 2024) discuss. In (Martiskova and Svec, 2020) researchers revealed that both genders are equally willing to deny a purchase, due to extreme personal data requirements. In (Mathur et al., 2019), the authors investigated the prevalence of deceptive design practices in 11K popular shopping websites, discovering that about 11.1% displayed at least one instance of dark patterns.

In (Pabian et al., 2020), authors identified key security threats related to payment methods, personal data, and purchased goods for both customers and sellers. Researchers in (Degutis et al., 2023) indicated that consumers value the expected give-and-take from

e-commerce providers more than the direct benefits of data disclosure. Authors in (Diaz et al., 2016) demonstrated that privacy threats are present in all stages of the e-shopping process, thus protecting only individual stages is insufficient. On top of that, in (di Vimercati et al., 2020), the authors discussed the challenges of balancing data availability for analysis with individual control over personal data. In (Morić et al., 2024), researchers emphasize the importance of robust data security measures in e-shops, presenting a framework that integrates legal, technological, and procedural elements to enhance data protection and consumer trust, aligned with standards like GDPR.

In addition, the impact of Secure Multiparty Computing (MPC) on traditional factors such as control, trust, and risk in data sharing decisions enhances control, reduces the need for interorganizational trust and prevents data leakage (Agahari et al., 2022). At the same time, MPC enables a "privacy-as-a-service" business model, enhancing security and reducing trust dependencies on data marketers, while providing new revenue opportunities through analytics and privacy services (Agahari et al., 2021). Researchers in (Sakalauskas and Kriksciuniene, 2024) introduced an algorithm, which uses clickstream data for targeted advertising to high-value customers. By measuring user activity, advertisers will improve ad performance, while costs can be reduced.

6 CONCLUSION & DISCUSSION

6.1 Summary

In this work, we explore user privacy breaches on e-shop platforms in a global scale. We find that e-commerce websites interact with the most third-party entities, suggesting that there is a potential leak of private information towards third parties. In fact, we study 200 distinct e-shops platforms from countries around the world and discover that nearly 30% of these leak at least one piece of sensitive information to a third-party entity. In addition, we find that Web conglomerates such as Facebook collect sensitive user information from multiple e-shops, and that they can use this information to match shopping habits with online user profiles. Finally, we highlight that even minimal interactions with these platforms can lead to substantial privacy risks, as a profile can be compromised after engaging with just five online stores. These findings highlight the need to take protective measures, enhance privacy protection and transparency in handling data over retail online shops.

6.2 Discussion

The findings of this work emphasize the need for improved transparency, privacy, and trust regarding personal data. While platforms likely disclose the sharing of sensitive information in their lengthy and obscure terms and conditions, there is a mismatch with user expectations when registering on online stores. Users typically assume their sensitive information will only be used for purchasing products, not shared with unknown third parties. Consumers are increasingly aware of privacy risks, which may influence their trust and shopping behavior in online shops.

6.3 Ethical Considerations

In this study, we made deliberate efforts to study the e-commerce ecosystem without disrupting it. The data collection process consisted of manual actions, minimizing the use of instrumented operations to a minimal. In Section 3, our automated system visited only the landing page of each website to assess third-party interactions, ensuring no impact on its performance. Each website was processed once, one at a time, simulating real user activity. Lastly, no personal data was collected or shared, adhering to research ethics principles (Rivers and Lewis, 2014).

REFERENCES

Agahari, W., Dolci, R., and de Reuver, G. (2021). Business model implications of privacy-preserving technologies in data marketplaces: The case of multi-party computation. In *29th European Conference on Information Systems (ECIS 2021) A Virtual AIS Conference: Human Values Crisis in a Digitizing World*, pages 1–16. Association of the Information Systems.

Agahari, W., Ofe, H., and de Reuver, M. (2022). It is not (only) about privacy: How multi-party computation redefines control, trust, and risk in data sharing. *Electronic markets*, 32(3):1577–1602.

Aqeel, W., Chandrasekaran, B., Feldmann, A., and Maggs, B. M. (2020). On landing and internal web pages: The strange case of jekyll and hyde in web performance measurement. In *Proceedings of the ACM Internet Measurement Conference*, page 680–695, New York, NY, USA. Association for Computing Machinery.

Bekos, P., Papadopoulos, P., Markatos, E. P., and Kourtellis, N. (2023). The hitchhiker's guide to facebook web tracking with invisible pixels and click ids. In *Proceedings of the ACM Web Conference 2023*, pages 2132–2143.

Broeder, P. (2020). Culture, privacy, and trust in e-commerce. *Marketing from Information to Decision Journal*, 3(1):14–26.

Center, P. R. (2024). Social media and news fact sheet. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/.

Degutis, M., Urbonavičius, S., Hollebeek, L. D., and Anselmsson, J. (2023). Consumers' willingness to disclose their personal data in e-commerce: A reciprocity-based social exchange perspective. *Journal of Retailing and Consumer Services*, 74:103385.

di Vimercati, S. D. C., Foresti, S., Livraga, G., and Samarati, P. (2020). Toward owners' control in digital data markets. *IEEE Systems Journal*, 15(1):1299–1306.

Diaz, J., Choi, S. G., Arroyo, D., Keromytis, A. D., Rodriguez, F. B., and Yung, M. (2016). Privacy threats in e-shopping. In *Data Privacy Management, and Security Assurance: 10th International Workshop, DPM 2015, and 4th International Workshop, QASA 2015, Vienna, Austria, September 21–22, 2015. Revised Selected Papers 10*, pages 217–225. Springer.

DuckDuckGo (2020). Duckduckgo tracker radar. https://github.com/duckduckgo/tracker-radar.

Fabbro, R. (2024). Visa will give customer data to retailers for ai-targeted ads. https://qz.com/visa-data-tokens-ai-marketing-retailers-transactions-1851481158.

Firstpost (2024). Amazon confirms security breach where employee data of millions was leaked. https://www.firstpost.com/tech/amazon-confirms-security-breach-where-employee-data-of-millions-was-leaked-13834672.html.

Flores, R., Perine, C., Remorin, L., and Reyes, R. (2022). Examining security risks in logistics apis used by online shopping platforms. https://www.trendmicro.com/vinfo/us/security/news/online-privacy/pii-leaks-and-other-risks-from-unsecure-e-commerce-apis.

Forbes (2024). 35 e-commerce statistics of 2024. https://www.forbes.com/advisor/business/ecommerce-statistics/.

Group, D. M. (2024). Dpg media belgium extends digital advertising offering with exclusive retail data from carrefour via unlimitail and tom&co. https://www.dpgmediagroup.com/extension-digital-advertising-offering-retaildata-carrefour-unlimitail-tomenco.

Gurung, A. and Raja, M. K. (2016). Online privacy and security concerns of consumers. *Information & Computer Security*, 24(4):348–371.

Klappholz, S. (2024). National public data breach: Lawsuit claims failed to protect billions of personal records. https://www.itpro.com/security/data-breaches/national-public-data-breach-lawsuit-claims-nearly-three-billion-people-had-personal-data-exposed.

Kurniawan, I. D. and Setyawan, V. P. (2024). The importance of protecting e-commerce consumer personal data. *IJOLARES: Indonesian Journal of Law Research*, 2(2):51–55.

Lakshmanan, R. (2025). European privacy group sues tiktok and aliexpress for illicit data transfers to china. https://thehackernews.com/2025/01/european-privacy-group-sues-tiktok-and.html.

Langone, A. (2019). At least 80% of shopping apps leak users' data. here's how to protect yourself. https://money.com/at-least-80-of-shopping-apps-leak-users-data-heres-how-to-protect-yourself/.

Martiskova, P. and Svec, R. (2020). Digital era and consumer behavior on the internet. In *Digital Age: Chances, Challenges and Future 7*, pages 92–100. Springer.

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., and Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11k shopping websites. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW).

Microsoft (2020). Playwright". https://playwright.dev/.

Morić, Z., Dakic, V., Djekic, D., and Regvart, D. (2024). Protection of personal data in the context of e-commerce. *Journal of cybersecurity and privacy*, 4(3):731–761.

Okeke, R. I., Shah, M. H., and Ahmed, R. (2013). Issues of privacy and trust in e-commerce: Exploring customers' perspective. *Journal of Basic and Applied Scientific Research*, 3(3):571–577.

Pabian, A., Pabian, B., and Reformat, B. (2020). E-customer security as a social value in the sphere of sustainability. *Sustainability*, 12(24):10590.

Pagey, R., Mannan, M., and Youssef, A. (2023). All your shops are belong to us: Security weaknesses in e-commerce platforms. In *Proceedings of the ACM Web Conference 2023*, page 2144–2154, New York, NY, USA. Association for Computing Machinery.

Papadogiannakis, E., Papadopoulos, P., Kourtellis, N., and Markatos, E. P. (2021). User tracking in the post-cookie era: How websites bypass gdpr consent to track users. In *Proceedings of the Web Conference 2021*, WWW '21, page 2130–2141, New York, NY, USA. Association for Computing Machinery.

Rauti, S., Carlsson, R., Mickelsson, S., Mäkilä, T., Heino, T., Pirjatanniemi, E., and Leppänen, V. (2024). Analyzing third-party data leaks on online pharmacy websites. *Health and Technology*, 14(2):375–392.

Rivers, C. M. and Lewis, B. L. (2014). Ethical research standards in a world of big. *F1000Research*, 3.

Sakalauskas, V. and Kriksciuniene, D. (2024). Personalized advertising in e-commerce: Using clickstream data to target high-value customers. *Algorithms*, 17(1):27.

Similarweb-LTD (2025). Similarweb digital intelligence. https://www.similarweb.com/.

Smith, B. (2024). Arkansas sues chinese online retailer temu, claims site illegally accessing user information. https://www.kark.com/news/state-news/arkansas-sues-chinese-online-retailer-temu-claims-site-illegally-accessing-user-information/.

Vlachogiannakis, I. (2025). Open-source data. https://github.com/gvlachogiannakis/e-shop-privacy-leaks.

Weigand, S. (2023). Over 12backups. https://www.scworld.com/news/over-12-of-online-stores-accidentally-leak-data-during-private-backups.