April 16, 2023 preliminary draft – please do not cite or forward without permission of the authors

Cookie Intermediaries: Does Competition Leads to More Privacy?

Arion Cheong, D. Daniel Sokol, and Tawei Wang[1]

This draft includes preliminary results. Comments are very welcomed.

Abstract

In this study, we investigate the voluntary sharing of customer information by firms with third-party organizations and the resulting privacy consequences. We collect first-party cookies of U.S. firms that provide online services and analyze how they share customer information through these cookies. Our result finds that partners with significant market power are less likely to share customer information with third-party organizations. We also report that market competition faced by partners has a greater impact on their data sharing policies than the market concentration of data brokers. Our analysis indicates that sharing customer information to data brokers with a lower market share increases the risk of customer privacy information leakage, while sharing customer information with registered data brokers that have a higher market share reduces the risk of customer privacy information leakage. The findings highlight the importance of considering market structure and competition in decision-making regarding privacy policy.

Keywords: cookie intermediary; privacy; consumer protection; market competition

JEL Classification: M21, M37, M38, M48,

[1] Arion Cheong is an Assistant Professor of Accounting at Stevens Institute of Technology. Email: acheong@stevens.edu. D. Daniel Sokol is the Carolyn Craig Franklin Chair in Law and Professor of Law and Business at the USC Gould School of Law and Marshall School of Business. Email: dsokol@usc.edu. Tawei (David)Wang is an Associate Dean and a Driehaus Fellow at the Driehaus College of Business and an Associate Professor at the School of Accountancy & MIS, DePaul University. Email: david.wang@depaul.edu. We thanks comments from presentations at Berkeley, Notre Dame, Monash, LSE, INFORMS, University of Florida, and USC.

April 16, 2023 preliminary draft – please do not cite or forward without permission of the authors

Cookie Intermediaries: Does Competition Leads to More Privacy?

1. Introduction

With the rise of digital transformation and digital marketing practices, data brokers that collect individual personal information and partners that share or resell their customer information plays a critical role in the digital economy. Since these data brokers do not have a direct relationship with customers, they obtain individual information from other partners that have first-hand customer information to build up comprehensive profiles of individuals. Concerns about data brokers in the press and by federal and state regulators suggest that the role of data brokers is a growing issue of concern for policymakers largely due to the low level of transparency and to govern data broker data collection and sharing practices (FTC 2014, Brill 2013).

Until recently, many data brokers have relied on an internet technology called third-party cookies that are created and placed by data brokers to collect information about the users that visit their partners' websites. However, this has changed recently. Browser platforms, including Google, recently announced that they are phasing out third-party cookies in their platforms, following other major tech firms, such as Apple and Mozilla. In response to change in using third-party cookies, data brokers are now asking their partners to collect their customer information themselves with first-party cookies and other data tracking technologies and transmit the information through alternative channels. For example, a recent study by Cookiebot (2019) reveals that Facebook had deployed a first-party cookie, "fbp_", to their partners' websites to collect user information and forwarded the information to their server for the Facebook Marketing Program. Also, the study shows that these first-party cookies were implemented on government websites for healthcare in

UK and Irish to capture personal information about HIV and mental illness through Facebook. Despite of the collection and sharing of sensitive personal information, these activities are not highly regulated.

In order to investigate this new phenomenon of data broker's new data collection and sharing model, we build on the growing literature that examines the interaction between market competition and privacy (Jia, Jin, & Wagman 2021; Johnson, Shriver, & Goldberg 2022; Sokol & Zhu 2021; Miller & Tucker 2009; Campbell, Goldfarb, & Tucker 2015, Raith 1996).

First, our study examines the effect of market concentration on customer information leakage, which still remains unclear. That is, firms may compete on the quality of privacy (Brough et al. 2022; Johnson, Shriver, & Du 2020; Tang, Hu, & Smith 2008), which becomes a non-price competition dimension.[2] Specifically, firms in a less concentrated market can consider privacy as a competitive advantage. On the other hand, prior studies emphasize the importance of understanding a firm's resources and capabilities, adopting a stakeholder-centric approach, and considering market structure and competitive dynamics when managing customer data. In line with this discussion, we assess how the market structure affects firms' data sharing activities with data brokers and how it relates to privacy information leakage.

Second, our study contributes to the literature in privacy and competition as well as the general literature on data brokers (Gu et al. 2021; West 2019; Braulin & Valleti 2016; Bélanger & Crossler 2011; Liu & Serfes 2006), and literature on data breaches (D'Arcy et al. 2020; Janakiraman, Lim, & Rishika 2018; Sen & Borle 2015) by identifying the data suppliers to the data brokers and how such data exchange activities may lead to privacy breaches. Our identification mechanism can

[2] We do so in ways that look at competition generally but not in the same way as antitrust-specific markets and examine competition tradeoffs with privacy.

benefit customers' privacy by continuously monitoring the data collection activities of such cookie intermediaries and their data exchange activities.

Data brokers have raised concerns about the lack of consumer privacy and data protection regulation. Stricter regulation of the data broker industry has been demanded, and antitrust actions and privacy regulations have been suggested. The FTC has recommended that data brokers be required to disclose their data collection practices and provide consumers with access to their data, and several states have enacted legislation that defines data brokers and requires them to report certain information. However, challenges in defining what constitutes a data broker and the complex nature of the industry make effective regulation difficult. Moreover, the effectiveness of regulation that balances consumer privacy and competition promotion is still unclear. The study aims to investigate how the effectiveness of data breach regulation is affected by market concentration level.

To address our research question, we focus on the data suppliers (i.e., partners) of data brokers, whom we label "cookie intermediaries." These cookie intermediaries provide their customers' information to data brokers through first-party cookies. We investigate every U.S. public firm's website and identify all first-party cookies that provide information to data brokers.[3] We then link these firms to the data brokers registered through the California Office of Attorney General. Then for each partner that shares information with data brokers, we calculate the corresponding industry product market competition index, which gives us a proxy to evaluate the competitiveness of a given market. Last, using a unique database provided by a major dark web monitoring firm that contains the dark web posts with Personally Identifiable Information (PII), we capture information

[3] Since partners themselves via first-party cookies collect customers' personal information, the partner firms do not require further notification to their customers beyond their privacy policies, which risks customer privacy (FTC 2013).

leakage from cookie intermediaries to the dark web. Our main findings suggest that partners in markets that are less concentrated share more customer information with third-party entities and experience more customer privacy information leakage than the partners in more highly concentrated markets.

We find that partners with significant market power are less likely to share customer information with third-party organizations that potentially due to their stronger reputation and less economic incentive to sell customer privacy information. The result shows that market competition faced by partners has a greater influence on their data sharing policies than the market concentration of data brokers. Further, we observe that sharing customer information with data brokers that have a lower market share increases the likelihood of customer privacy information leakage. On the other hand, sharing customer information with registered data brokers with a higher market share reduces the likelihood of customer privacy information leakage, indicating the effectiveness of established policies and procedures of registered data brokers to manage privacy risks.

In the remaining sections of this study, we provide a summary of relevant literature and our hypotheses. Then, we present our preliminary analysis and the results. We conclude with the policy implications of our results and identify areas for future research.

2. Privacy, Market Concentration, Customer Data Sharing

First-Party Data Holder and Data Collection

Data brokers usually purchase data from a first-party data holder, e.g., Amazon, Facebook, or Bank of America, that has a direct relationship with their end customers. For example, Bank of

America specifies on their consumer privacy notice website that they share personal information with their affiliates and even nonaffiliates for marketing purposes, which a consumer could not limit the sharing of once using Bank of America's service.[4]

Compared to the first-party data holder, data brokers operate in the upstream market to collect information from a wide variety of sources (i.e., first-party data holders) and process it into a structured form to extract insight and information. The data collection includes public records about individuals and even data sold or licensed by the first-party data holders. the major portion of the information collected by the data brokers is used for marketing products to classify customers and generally place them into a "bucket" based on their individual attributes learned by the information (FTC 2014).

Table 1 provides examples of the information collected by major data brokers (Google and Oracle) and the benefits they provide to their partners through customer information sharing. The benefits can be categorized into two main groups: the advertisement revenue multiplier model and the cross-channel marketing model. In the advertisement revenue multiplier model, websites that provide advertising space can earn higher Cost-per-Click (CPC) by sharing their customer information with marketers. In the cross-channel marketing model, partners can upload customer information to the online platform and data brokers will provide matching customer information from their database. Data brokers take ownership and control over the aggregated information uploaded to their platform.

[Insert Table 1 Here]

[4] Refer to the website for further details about Bank of America's consumer privacy notice.

Third-party cookies have been widely used by data brokers, especially advertisers, to collect information about the users' online activities across different websites and devices. However, Google has announced that it will block the usage of any third-party cookies by 2023. This follows changes made by Mozilla (Firefox) and Apple (Safari) toward their browsers that limit third party collection of information. Although the demise of cookies seemed to be the end of the online user tracking, data brokers recently developed an efficient solution to circumvent the anti-tracking mechanism: asking first party partner firms to collect the user information by themselves by implementing first-party cookies. First-party cookies are also created and deployed by the first party partner website to track user behavior while storing all the collected information in the user's computer.

For instance, when Safari's anti-tracking mechanism blocked third-party cookies, Facebook developed a first-party cookie called "_fbp" and shared it with their partner websites until recently. Partners were permitted to use the first-party cookie supplied by Facebook, which gathered user information and tagged it with a unique identifier. The information collected by the cookie was then sent to Facebook through a pixel tracker embedded on the website, as shown in Figure 1.

[Insert Figure 1 Here]

The sharing of customer data to data brokers can lead to serious privacy breaches. For instance, the Facebook case highlights how first-party partners can collect user information through first-party cookies and share it with data brokers for marketing purposes. This data can then be sold to third parties or used for targeted advertising. If this data falls into the wrong hands, it can be misused and exploited for malicious purposes, potentially putting consumers' personal information and privacy at risk.

Customer Information Leakage and Market Concentration

In recent years, firms are increasingly collecting and sharing customer information with third parties in order to gain a competitive advantage. However, this practice has raised concerns regarding privacy and security. Our study adds to the existing literature on market structure and privacy by investigating how firms share customer information voluntarily and the potential consequences of such actions.

The strategic management literature emphasizes the importance of understanding a firm's resources and capabilities when making decisions about managing customer information. Resource-Based Theory suggests that customer information is a valuable resource that contributes to a firm's competitive advantage (Bharadwaj & Soni 2020; Krsnova, Spiekermann, Koroleva, & Hildebrand 2010; Li & Ye 2019; Smith, Dinev, & Xu 2011). As a result, it is crucial for firms to understand their resources and capabilities when making decisions about privacy policies. In addition, stakeholder theory emphasizes the need to consider the interests of all stakeholders, particularly customers, when developing privacy policies. Therefore, a stakeholder-centric approach to managing customer data is essential, which involves balancing the interests of all stakeholders (Culan & Bies 2003; Gao & Chen 2021; Smith, Milberg, & Burke 1996; Zhang & Benbasat 2004).

The management literature highlights that market structure affects the behavior of firms and the outcomes in customer information protection. Early study by Porter (1980) suggests that understanding market structure and competitive dynamics is critical in determining the optimal business strategy for the firm. The literature provides mixed results on whether market competition provides positive or negative privacy outcomes. For instance, Chen and John (2018) found that consumers are more likely to share personal information with firms in less competitive markets, whereas Singh and Thambusamy (2021) found that increased competition can reduce privacy concerns by providing consumers with more choice and bargaining power.

Overall, the literature indicates that firms need to understand their resources and capabilities, adopt a stakeholder-centric approach, and consider market structure and competitive dynamics when managing customer data. However, the effect of market concentration on customer information leakage and data sharing activities remains an open question. To the best of our knowledge, no studies provide a clear answer to this question. Our study contributes to the literature by examining how market structure affects customer information leakage and its potential implications for privacy protection.

Data Brokers and Regulation

The rise of data brokers as a major player in the data economy has sparked worries about safeguarding consumer privacy and data protection. These concerns stem from the lack of transparency in data broker operations and the potential abuse of consumer data. Therefore, there are growing demands for stricter regulation of the data broker industry. Additionally, research indicates that data brokers are susceptible to data breaches, which can lead to the sale of consumers' personal information on the dark web (Ponemon Institute, 2014), highlighting the need for enhanced regulation to ensure the protection of consumer data.

Accordingly, market regulation such as antitrust enforcement and privacy have received attention, but opinions on their effectiveness are mixed. Antitrust actions may prevent firms from acquiring data that could be used to enhance their market power and reduce competition, thus promoting privacy (Wu & Stucke, 2018; Rai & Boyle 2018). However, concerns have been raised that antitrust enforcement may have unintended consequences for privacy, such as leading to the breakup of large firms that have better privacy protection measures (Irion & De Hert 2018). Additionally, antitrust enforcement may result in the emergence of smaller, less-resourced firms that are more likely to engage in data breaches or other privacy violations (Rai & Boyle 2018).

To address this issue, the Federal Trade Commission (FTC) has recommended that data brokers be required to disclose their data collection practices and provide consumers with access to their data (FTC 2014). In a more recent report to Congress in 2021, the FTC advocates for legislation to grant them more regulatory powers regarding privacy and data security.

Also, two states have enacted legislation that defines data brokers and requires them to report certain information. Vermont released Data Broker Regulations in 2018 (9 VSA § 2430) that mandate data brokers to register with the Secretary of State annually. California also requires data brokers to publish their information on their websites. In Vermont, a data broker is defined as a business or unit that knowingly collects, sells, or licenses the personal information of a customer with whom the business does not have a direct relationship. In California, a data broker is defined as a business that knowingly collects and sells the personal information of a consumer with whom the business does not have a direct relationship (Cal. Civ. Code § 1798.99.80).

Regulating data brokers is a challenging task due to the lack of a clear definition of what constitutes a data broker, as well as the complex nature of the data broker industry. This complexity makes it difficult to create effective regulations to protect consumer privacy. For example, the lack of enforcement and low penalties for non-registration contribute to the challenge of regulating data brokers. Furthermore, under the California data broker law, registered data brokers are required to provide a pre-collection notice before selling or sharing any collected personal information. However, registration remains a strategic means for data brokers to obtain the pre-collection notice exemption, which in turn increases the likelihood of increasing data sharing activities and potential customer privacy information leakage.

In conclusion, the literature suggests that data brokers have the potential to pose significant risks to consumer privacy and data protection. Regulation is necessary to ensure that data brokers are transparent in their operations and take appropriate measures to protect consumer data. However, there are challenges in understanding the effectiveness of regulation that strikes a balance between protecting consumer privacy and promoting competition. In our study we will study how the effectiveness of data breach regulation considering the market concentration level.

3. Hypothesis Development

Market Concentration and Customer Information Leakage (Partner-level)

Building on prior literature regarding competition and firm performance, we explore how market structure (proxied by market concentration level) influences partners' data collection and sharing practices. That is, given the competition faced by a partner, each partner makes a strategic decision between customer privacy or sharing customer information.

Considering the strategic decision made by the partners related to their consumer privacy, we empirically examine two main research questions to study the association between the market structure of cookie intermediaries and customers' privacy breaches. First, we examine whether the first party partner firm in a concentrated industry provides less customer information to data brokers. The following hypothesis summarizes our first research question.

H1a: Partners in a more concentrated industry share more customer information with data brokers than partners that are in a less concentrated industry.

Related to the first research question, our second research question examines whether partners in more concentrated markets experience fewer privacy breaches. Further, we compare the consequences of data sharing to data brokers across different levels of competition.

H1b: Partners in a more concentrated market experience more customer information leakage when sharing customer information with third-party data brokers.

Market Concentration and Customer Information Leakage (Data Broker-level)

In addition to our analysis of the effect of partner-level market concentration on privacy, we examine whether market competition at the level of data brokers can affect the level of consumer privacy information leakage. Utilizing our unique data source obtained from the largest online code repository, *GitHub*, we extend our first hypothesis by considering market share among data brokers. Specifically, we empirically examine whether first party partner firms share more information with data brokers with higher market share in the data broker industry.

H2a: Partners in a more concentrated industry share more information with data brokers with lower market share.

Next, we extend our second hypothesis to study how data broker-level market competition affects consumer privacy leakage. Accordingly, we examine how data broker market competition may affect the level of consumer privacy information leakage.

H2b: Partners sharing information with a data broker with lower market share experience more privacy breaches.

Effectiveness of Data Broker Regulation

We expect first-party data holders are more likely to share customer information with registered data brokers compared to unregistered data brokers. This hypothesis is based on the premise that registered data brokers are required to disclose their data collection practices and provide consumers with access to their data. In contrast, unregistered data brokers may lack such transparency and accountability measures, making them less attractive partners for firms seeking to protect their customers' privacy. As such, we assume a positive relationship between data broker registration status and the amount of customer information shared by partners. This hypothesis will be tested empirically by examining the data sharing practices of firms with registered and unregistered data brokers.

H3a: Partners in less concentrated industry share more customer information with registered data brokers than unregistered data brokers.

The next hypothesis suggests that the market share of registered data brokers can affect the amount of customer information leakage. Partners that share customer information with registered data brokers that have a higher market share are likely to experience less customer information leakage than those that share information with non-registered data brokers. This may be because registered data brokers with a higher market share are more established and have stronger data protection measures in place (i.e., Resource-based Theory), making them less vulnerable to data breaches. Additionally, higher market share data brokers may have more resources to invest in security and compliance measures, making them a more reliable and trustworthy partner for firms seeking to protect their customers' privacy. Ultimately, this hypothesis suggests that the market share of registered data brokers can have a significant impact on the effectiveness of customer data protection measures employed by firms.

H3a: Partners sharing customer information with registered data brokers with higher market share experience less customer information leakage.

In sum, Figure 2 summarizes the hypotheses developed in our study regarding the relationship between market concentration, data sharing activities, customer privacy information leakage, and data broker regulation.

[Insert Figure 2 Here]

4. Data and Measurement

In our empirical analysis, we utilize several unique data sources. Figure 3 illustrates our data collection and variable measurement process.

[Insert Figure 3 Here]

First, we identify 9,161 U.S. public firms where their websites are captured and stored by *Internet Archive*. It allows to observe previous renditions of websites and retrieve the initial source code of websites that are not directly accessible anymore. We exclude records with source codes that are smaller than 1,000 words or more than 200,000 words, which are mostly caused by broken HTML elements. Next, we exclude firms that does not have any financial information provided by Compustat and market concentration measure by Hoberg and Philips (2016). Our final sample includes 575 unique firms that results in 2,832 firm-year observation.

Next, we retrieve the list of first-party marketing cookies and the data brokers that controls information collected by those cookies from *CookieDatabase.org*. Each cookie that a user encounters on a website can provide information to third-party organizations that are listed in *CookieDatabase.org*. These organizations are commonly referred to as data brokers, who are deemed as "data controllers" due to their exclusive control over the information collected from users through first-party cookies. The data collected from first-party cookies is often encrypted using techniques such as hashing before being sent to the data brokers' servers via secure channels. In this study, the third-party organizations identified in *CookieDatabase.org* are classified as data brokers (both registered and unregistered).

We count the number of first-party marketing and tracking cookies that are implemented on each firm's website, which is labeled as *DS*. From our list, we identify most of the advertisement firms (e.g., Kentico) and e-commerce data analytic service providers (e.g., Shopify). Also, we include major managed security service providers (e.g., Imperva) in our list. These managed security service providers also collect the device and user information from their partners' websites. For example, Imperva states in their privacy policy statement that: "These cookies gather information about your browsing habits and are used to help us understand your interests so we can deliver content that is more relevant to you. For example, they may be used to deliver targeted advertising or to limit the number of times you see an advertisement. They also help us measure the effectiveness of advertising campaigns on our Digital Properties. We may share this information with other parties, including our advertisers and other service providers."[5]

Customer Privacy Breaches

In our study, we perform an analysis to examine the customer privacy information leakage (*CPL*) that occurs due to data sharing. To measure *CPL,* we search public sources for instances of customer information breaches. These public sources included well-known breach notification websites such as *HaveIbeenPwned.com*, the U.S. Department of Health and Human Services, and the State Office of the Attorney General in Washington, Oregon, and California. We also utilized

[5] Refer to this website of Imperva to obtain the original privacy policy statements on cookies: https://www.imperva.com/trust-center/cookie-notice/

the Cybersecurity dataset from *Audit Analytics*. Through this process, we identified a total of 617 breaches within our sample.

Market Competition between Industry and Data Brokers

In order to measure the level of market concentration in different industries, we utilize Hoberg and Philips' (2016) Text-based Network Industry Concentration (TNIC) Data, which provides the Herfindahl-Hirschman Index (*HHI*) for each industry. This measure is a more current and relevant measure of industry concentration as it is updated yearly based on the business description disclosed in Item 1 of the 10-K filing. The resulting measure of market concentration, labeled *MC_Firm*, ranges from 0 to 1 in the sample, with a median of 0.195. Descriptive statistics are presented in Table 3. The TNIC-HHI measure is used to proxy the level of competition faced by the firms in the respective industries.

[Insert Table 3 Here]

To estimate the market concentration level of data brokers, we suggest the measure of market share of each data broker, which is referred to as *MC_DB*. We count the number of code repositories on *Github* that mention the name of data brokers in the *Topic* tag to calculate this measure. *Github* is a leading platform where developers and companies create, develop and maintain software. We use the Search API provided by *Github* to identify the number of topics that mention the name of the data broker in both public and private repositories. *MC_DB* measures the relative market share of data brokers, where a higher value of *MC_DB* indicates a larger market share of data brokers in the data broker industry. The study found that *MC_DB* varies between 0 to 1 in their sample, with a median of 0.599.

In addition to our main effect variables, we consider firm-specific variables, such as the logarithm amount of Net Income (*NI*) and Total Assets (*AT*), performance measures such as Return on Assets (*ROA*) and Return on Equity (*ROE*), negative income (*Neg*), and the age of the firm (*Age*). We selected to control the firm-specific effects following the prior literature on cybersecurity and data breaches (e.g., Wang et al. 2013; Gordon et al. 2010).

5. Analysis and Results

The Effect of Market Concentration on Consumer Privacy Leakage

The first research question aims to determine whether partners in less concentrated industries share more information with data brokers than those in more concentrated industries. Our second research question is related to the first and investigates whether firms in less concentrated markets are more likely to experience privacy breaches when they share data through first-party cookies.

$$DS_t = MC\_Firm_t + Control\ Variables_t + Fixed\ Effects_t \tag{1}$$

$$CPL_t = DS_t \times MC\_Firm_t + Control\ Variables_t + Fixed\ Effects_t \tag{2}$$

In Table 4, Column 1 presents the results of a regression analysis conducted to test the first hypothesis, which examines whether the level of market competition faced by partners would have an effect on their use of trackable first-party cookies. The coefficient of the market competition effect in the regression analysis was found to be negative and statistically significant, with a *t*-statistic of -2.045. The findings of the study indicate that partners facing greater competition (as measured by a proxy of lower concentration) tend to use more trackable first-party cookies than those facing less competition (as measured by a proxy of higher concentration).

[Insert Table 4 Here]

This suggests that partners with significant market power are less likely to share customer information with third-party organizations. Specifically, our results demonstrate that partners facing higher competition tend to use more trackable first-party cookies and share less customer information with third-party organizations, which highlights the importance of considering market structure in understanding the privacy practices.

Hypothesis 1b investigates the potential interaction effect between a partner's data-sharing activities and the level of market competition they face on the amount of customer information leakage. To test this hypothesis, we conducted a logit regression analysis, and the results of this analysis are presented in Column 3 of Table 4. Contrary to our initial expectations, our analysis indicates that there is no significant interaction effect between a partner's data-sharing activities and the level of market competition they face. This suggests that the level of competition faced by partners does not appear to have a significant impact on the relationship between data-sharing activities and customer information leakage.

The Effect of Data Broker Level Competition on Consumer Privacy Leakage

In this section, we examine the effect of data broker-level market share on the data sharing activities of firms and the resulting impact on consumer privacy information leakage. To test this relationship, we extended the OLS regression model specified in equations 1 and 2 by adding an interaction term with the market share of data brokers, as outlined in hypotheses provided in equations 3 and 4. Specifically, we hypothesized that firms with higher levels of data sharing activities with data brokers with lower market share would experience greater consumer privacy information leakage.

$$DS_t = MC\_DB_t \times MC\_Firm_t + Control\ Variables_t + Fixed\ Effects_t \tag{3}$$

$$CPL_t = DS_t \times MC_{DB_t} + \ Control\ Variables_t + Fixed\ Effects_t \qquad (4)$$

The results presented in Column 2 of Table 4 reveals a negative and statistically significant coefficient for data partner-level market competition (*MC_Firm*), indicating that partners facing higher market competition tend to implement more lenient data sharing policies (*t*-statistic: -2.663). In addition, we found a positive but moderate interaction effect between data broker-level market concentration (*MC_DB*) and data partner-level market competition (*MC_Firm*), as indicated by a *t*-statistic of 1.664. However, the coefficient for *MC_DB* is not statistically significant.

These findings from our study suggest that the level of market competition faced by partners has a greater influence on their data sharing policies than the market concentration of data brokers who they share customer information with. This shows that partners tend to exercise more caution in their data sharing practices when they face greater market power, possibly by investing more resources to protect customer privacy information in order to maintain their reputation among customers.

The results from a logit regression model in Column 4 of Table 4 reveal a negative and significant coefficient for the interaction between data broker-level market concentration (*MC_DB* × *DS*), with a *t*-statistic of -2.720. This suggests that the partners sharing more customer information with data brokers experience significantly higher customer privacy information leakage. Instead, the market concentration level of the data brokers plays a more critical role in determining the risk of customer privacy information leakage. In other words, the results suggest that partners are more likely to experience customer privacy information leakage when they share information with data brokers that have a lower market share, rather than when they share information with data brokers that has more market share.

Data Broker Registration and Customer Privacy Information Leakage

Equation 5 represents a regression model that examines the relationship between the data sharing activities (*DS*) of partners and the market competition level faced by partners (*MC_Firm*) while taking into account the data broker registration status of the data brokers that they share their customer information with. Equation 6 examines the relationship between the level of market competition faced by data brokers (*MC_DB*) and the registration status of data brokers (*DB_Reg*), and how they impact customer privacy information leakage (*CPL*).

$$DS_t = MC\_DB_t \times MC\_Firm_t \times DB\_Reg_t + Control\ Variables_t + Fixed\ Effects_t \tag{5}$$

$$CPL_t = DS_t \times MC\_DB_t \times DB\_Reg_t + \ Control\ Variables_t + Fixed\ Effects_t \tag{6}$$

In Table 5, we report the relationship between data sharing activities and customer privacy information leakage, taking into account the registration status of the data brokers that partners share customer information with. The results show that *DB_Reg* is statistically significant in column 1, with a t-statistic of 2.128. This suggests that registered data brokers are more likely to receive customer information from partners. However, the interaction term between *DB_Reg* and *MC_Firm* is statistically insignificant, indicating that partners share more customer information with registered data brokers regardless of their market concentration level. The findings suggest that data brokers that are registered in the data broker registry are preferred by partners for data sharing activities.

The interaction effect of the variables *DB_Reg*, *MC_DB*, and *DS* on customer privacy information leakage (*DB_Reg* × *MC_DB* × *DS*) is presented in column 2 of Table 5. The coefficient of the interaction term is -4.257, and it is statistically significant with a *t*-statistic of -10.21. This result suggests that sharing customer information with registered data brokers with higher market share reduces the likelihood of customer privacy information leakage. Therefore, we report that firms have less risk of customer privacy information leakage during their data sharing practices when they share customer information with registered data brokers with a higher market share.

Based on the results of our study, we find that registered data brokers are more likely to receive customer information from partners for data sharing activities, regardless of their market concentration level. Additionally, sharing customer information with registered data brokers that have a higher market share reduces the likelihood of customer privacy information leakage. These findings are summarized in Figure 4.

6. Additional Analysis

Data Sharing Pattern

It is crucial to consider similar data sharing patterns when predicting privacy risk because the patterns can differ across industries, types of data, and purposes of data sharing. The European Union Agency for Network and Information Security (ENISA) noted in a study that "different sectors have different data sharing patterns and requirements, which require different security measures to protect personal data" (ENISA 2018). Therefore, to assess privacy risks related to data sharing, it is necessary to cluster partners based on the data brokers they share customer information with.

This process involves generating a vector in which each element in the data broker list is recorded as either 1 or 0, indicating whether the partner shares customer information with a particular data broker. Clustering methods are then adopted to group entities that share similar data sharing patterns. The Within-Cluster Sum of Squares (WCSS) is a common evaluation metric used to determine the optimal number of clusters. In Table 6, WCSS is documented for each number of clusters, and based on the marginal changes in WCSS, it is determined that there are four clusters in the sample. Panel B shows the results of clustering the data into 4 groups or clusters. Additionally, a categorical variable called *Market Concentration* is generated, which is labeled as high when the level of market concentration of the partner or data broker is higher or the same as the median and is labeled as low when the level of market concentration is lower than the median.

Table 7 provides the ANOVA results for examining the relationship between data sharing activities (*DS*) and market concentration at the firm and data broker level. Panel A of the table shows that both models have a significant overall effect on data sharing activities. In Panel B of Table 7, the analysis is further broken down by clusters. The results indicate that partners with lower market concentration tend to share more customer information with data brokers, which aligns with the findings of Hypothesis H1a and H1b. Moreover, this implies that the relationship between data sharing activities and market concentration is not uniform across all industries and that different industries have unique patterns of data sharing. The findings also imply that different industries have unique patterns of data sharing, which highlights the need for industry-specific approaches to privacy risk assessment and management.

Panel A of Table 8 indicates that, even though data sharing patterns differ across each cluster of partners, the level of customer privacy information leakage increases when firms share their customer privacy information with data brokers that have lower market share. The *F*-statistic of 7.650 confirms that this result is statistically significant, which confirms that our results on Hypotheses 2a and 2b hold true. The finding that customer privacy information leakage primarily depends on the market concentration level of data brokers suggests that more breaches are occurring from low market share data brokers. This highlights the importance of carefully selecting data brokers with to minimize the risk of privacy breaches.

Cross Sectional Analysis with Dark Web

We conduct an additional cross-sectional analysis to measure customer privacy information leakage (referred to as *CPL_DW*) by searching for Personally Identifiable Information (PII) of customers related to each partner in the darknet market. The analysis addresses the concern that major cybersecurity breaches for large companies are more likely to be announced and receive public attention, which may create a selection bias. This bias may overlook the potential risks faced by small and medium-sized companies that may not have the resources to detect and address cybersecurity threats as effectively as larger companies. To address this concern, our analysis focus on gathering data on cybersecurity breaches across different types and sizes of companies, rather than relying solely on publicly reported incidents.

To do so, we utilize a dataset of 297,935 darknet market posts from 714 websites collected by a major dark web monitoring service provider, *DarkOwl*, between February 1 and July 31, 2020. We filter out the posts that have been detected to contain personally identifiable information or malicious keywords. We then use textual analysis to identify posts that mention the name of the partner in the title and contain user information such as email ID and password. For instance, we count the number of emails leaked along with passwords from posts that include the company name in the title. We count each unique email address with a password as one instance of consumer privacy leakage (CPL_DW). Our analysis resulted in identifying the number of instances of CPL_DW for each partner in our sample. From column 5 and 6 in Table 4, we conduct the same analysis on our cross-sectional dataset obtained from the darkweb. We find all the results are consistent with our panel analysis.

Consistent to our earlier finding, partners that share information with less competitive data brokers may be at a higher risk of customer privacy breach in the dark web. Overall, the finding highlights the importance of companies being cautious when sharing customer information with third-party data brokers, particularly smaller ones, and ensuring that appropriate data security measures are in place.

7. Conclusion

In today's digital age, data has become a valuable commodity, and first-party data holders often share this data with data brokers to generate revenue through targeted advertising. However, the sharing of customer information raises concerns about privacy breaches and the tradeoff between customer privacy and revenue generation. Our study aims to shed light on this issue by examining the impact of competition on data sharing activities of first-party data holders with data brokers and the occurrence of consumer privacy breaches.

Our findings suggest that partners with significant market power are less likely to share customer information with third-party organizations. This could be because these partners have a stronger reputation to maintain and less economic incentive to sell customer privacy information. Moreover, we find that the level of market competition faced by partners has a greater influence on their data sharing policies than the market concentration of data brokers. In other words, when partners face higher competition, they tend to use more trackable first-party cookies and share less customer information with third-party organizations. This indicates that market structure is an essential factor to consider in understanding privacy practices.

Additionally, our study shows that the market concentration level of data brokers plays a critical role in determining the risk of customer privacy information leakage. Specifically, sharing customer information with data brokers that have a lower market share increases the likelihood of customer privacy information leakage. This highlights the need for partners to exercise caution when sharing customer information with data brokers and to consider the market concentration level of the data brokers they are sharing information with.

Finally, we find that sharing customer information with registered data brokers with a higher market share reduces the likelihood of customer privacy information leakage. This suggests that registered data brokers have established policies and procedures to protect customer privacy information, and their higher market share may be indicative of their ability to effectively manage privacy risks.

As with any study, our study has several limitations that must be considered. Our results are based on the data obtained from an online code repository, breach reports, and a dark web monitoring firm. Although our results are limited to the available data sources, we find our result to be robust and consistent.

In summary, our study provides important insights into the impact of competition on data sharing activities and the occurrence of consumer privacy breaches. The findings highlight the need for industry-specific approaches to privacy risk management and emphasize the importance of considering market structure and competition when assessing privacy risks associated with data sharing.

References

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. MIS Quarterly, 35(4), 1017-1040.

Bharadwaj, A., & Soni, D. (2020). Impact of strategic customer relationship management on firm performance: The role of customer information management capabilities. Journal of Business Research, 113, 108-119.

Braulin, J., & Valleti, M. R. (2016). The determinants and welfare implications of data breaches. Journal of Public Economic Theory, 18(6), 852-867.

Brill, R. J. (2013). Reclaim your name. The New York Times. Retrieved from https://www.nytimes.com/2013/06/16/opinion/sunday/reclaiming-your-name-from-data-brokers.html

Brough, A. R., Chernev, A., & Sezer, O. (2022). Opting out of privacy? Consequences for consumer behavior and competition. Journal of Consumer Research, 49(3), 355-378.

California Civil Code, Division 3, Part 4, Title 1.81.5, Chapter 2, Section 1798.99.80 (2018)

Campbell, M. C., Goldfarb, A., & Tucker, C. (2015). Privacy regulation and online advertising. Management Science, 61(11), 2727-2744.

Chen, Y., & John, G. (2018). Consumer privacy and marketing in oligopolistic competition. Journal of Marketing Research, 55(1), 91-106.

Cookiebot. (2019). Tracking the trackers: Cookie syncing and the state of cross-device tracking. Retrieved from https://www.cookiebot.com/media/1309/cookiebot_cookie-syncing-report-2019.pdf

Culan, L., & Bies, R. J. (2003). Handbook of organizational justice. Routledge.

D'Arcy, J., Hovav, A., & Galletta, D. F. (2020). User awareness of data breaches: An empirical exploration of the role of phishing emails. Computers & Security, 94, 101854.

Federal Trade Commission. (2014). Data brokers: A call for transparency and accountability. Retrieved from https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Gao, H., & Chen, Y. (2021). Understanding customer privacy concerns: An examination of the privacy calculus model in e-commerce. Journal of Business Research, 125, 205-219.

Gordon, L. A., Loeb, M. P., & Lucyshyn, W. (2010). Sharing information on cybersecurity and business resilience: An experimental analysis. Journal of Public Economics, 94(11-12), 935-945.

Gu, B., Konana, P., Liang, X., & Zhang, Z. (2021). Information privacy in digital markets: A multidisciplinary research review. Journal of Management Information Systems, 38(1), 7-43.

Hoberg, G., & Phillips, G. (2016). Text-Based Network Industry Classifications: A Versatile Framework for Research and Policymaking. Journal of Economic Perspectives, 30(4), 59-88. doi: 10.1257/jep.30.4.59

Irion, K., & De Hert, P. (2018). The EU General Data Protection Regulation: Toward a new regime of global data governance. Columbia Journal of European Law, 24(2), 149-182.

Janakiraman, R., Lim, S., & Rishika, R. (2018). The role of financial incentives and data breach severity in consumers' security breach notification responses. Journal of Management Information Systems, 35(4), 999-1032.

Jia, P., Jin, G. Z., & Wagman, L. (2021). Privacy, market competition, and innovation. Management Science, 67(9), 5987-6006.

Johnson, J. P., Shriver, S. K., & Du, R. Y. (2020). Do consumers value data privacy? Evidence from a willingness to pay for privacy-protecting technologies. Journal of Marketing Research, 57(6), 1116-1132.

Johnson, J. P., Shriver, S. K., & Goldberg, L. R. (2022). The impact of competition on consumer privacy: Evidence from the adoption of HTTPS encryption. Journal of Public Policy & Marketing, 41(1), 122-137.

Krsnova, L., Spiekermann, S., Koroleva, K., & Hildebrand, C. (2010). Managing privacy in online social networks: The privileged position of business networks. Journal of Business Research, 63(9-10), 1101-1108.

Li, X., & Ye, Q. (2019). The value of personal information: Evidence from online behavioral advertising. Information Systems Research, 30(1), 211-234.

Liu, Q., & Serfes, K. (2006). On the efficiency of data brokers. The Journal of Industrial Economics, 54(4), 527-550.

Miller, M. K., & Tucker, C. (2009). Privacy protection, personalized marketing, and market structure. Journal of Marketing Research, 46(5), 552-562.

Ponemon Institute. (2014). The economics of the dark web.

Porter, M. E. (1980). Competitive strategy: Techniques for analyzing industries and competitors. New York: Free Press.

Rai, A. K., & Boyle, J. P. (2018). Antitrust remedies for labor market power. In The Future of Labor and Employment Law in the United States (pp. 99-120). Springer, Cham.

Raith, M. G. (1996). Competition and privacy. Antitrust Bulletin, 41(3), 725-748.

Sen, R., & Borle, S. (2015). The effect of data breach announcements beyond the stock price: Empirical evidence on market activity. Journal of Management Information Systems, 32(4), 84-116.

Singh, A. K., & Thambusamy, R. (2021). Market competition and consumers' online privacy concerns: An empirical study. Journal of Retailing and Consumer Services, 61, 102566.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. MIS Quarterly, 35(4), 989-1016.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices. MIS Quarterly, 20(2), 167-196.

Sokol, D. D., & Zhu, F. (2021). Competition, privacy, and personal data. Antitrust Law Journal, 86(2), 351-374.

Tang, Z., Hu, Y. J., & Smith, M. D. (2008). The effects of website design on purchase intention in online shopping: The mediating role of trust and the moderating role of culture. International Journal of Electronic Commerce, 13(1), 49-74.

Vermont Statutes Annotated, Title 9, Chapter 62, Section 2430 (2018)

Wang, T., Chen, C., & Germain, R. (2013). The Impact of Information Security Incidents on Firm Performance: An Empirical Investigation. Decision Support Systems, 56, 86-96.

West, D. M. (2019). Data brokers and the federal government: A new era of regulation. Brookings Institution.

Wu, T., & Stucke, M. (2018). Competition, innovation, and privacy on the internet: Economic analysis of Google's conduct in Europe. Journal of Competition Law & Economics, 14(1), 1-43.

Zhang, P., & Benbasat, I. (2004). Consumer trust in B2C e-commerce and the importance of social presence: experiments in e-products and e-services. Omega, 32(6), 407-424.

Figure 1: Circumvention technique to tracker blocking (Cookiebot 2019)

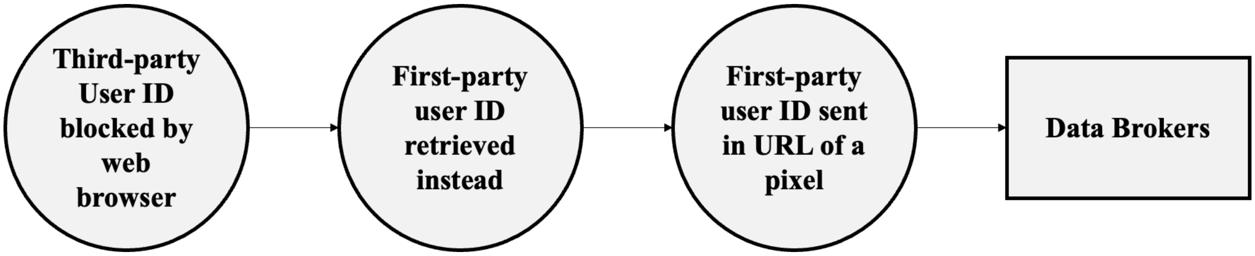

Third-party User ID blocked by web browser
First-party user ID retrieved instead
First-party user ID sent in URL of a pixel
Data Brokers

Figure 2. Hypothesis Summary

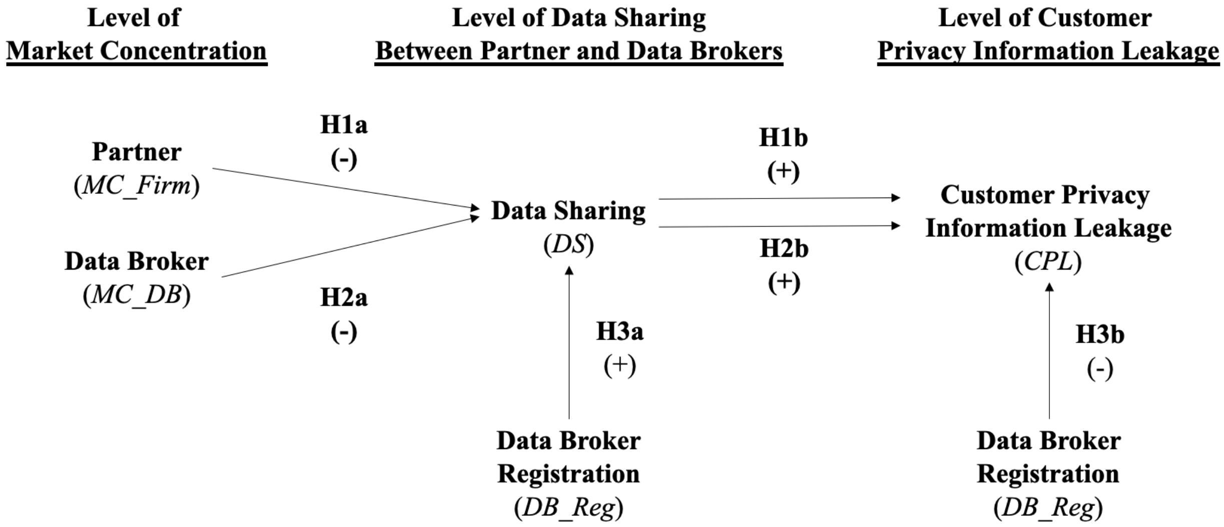
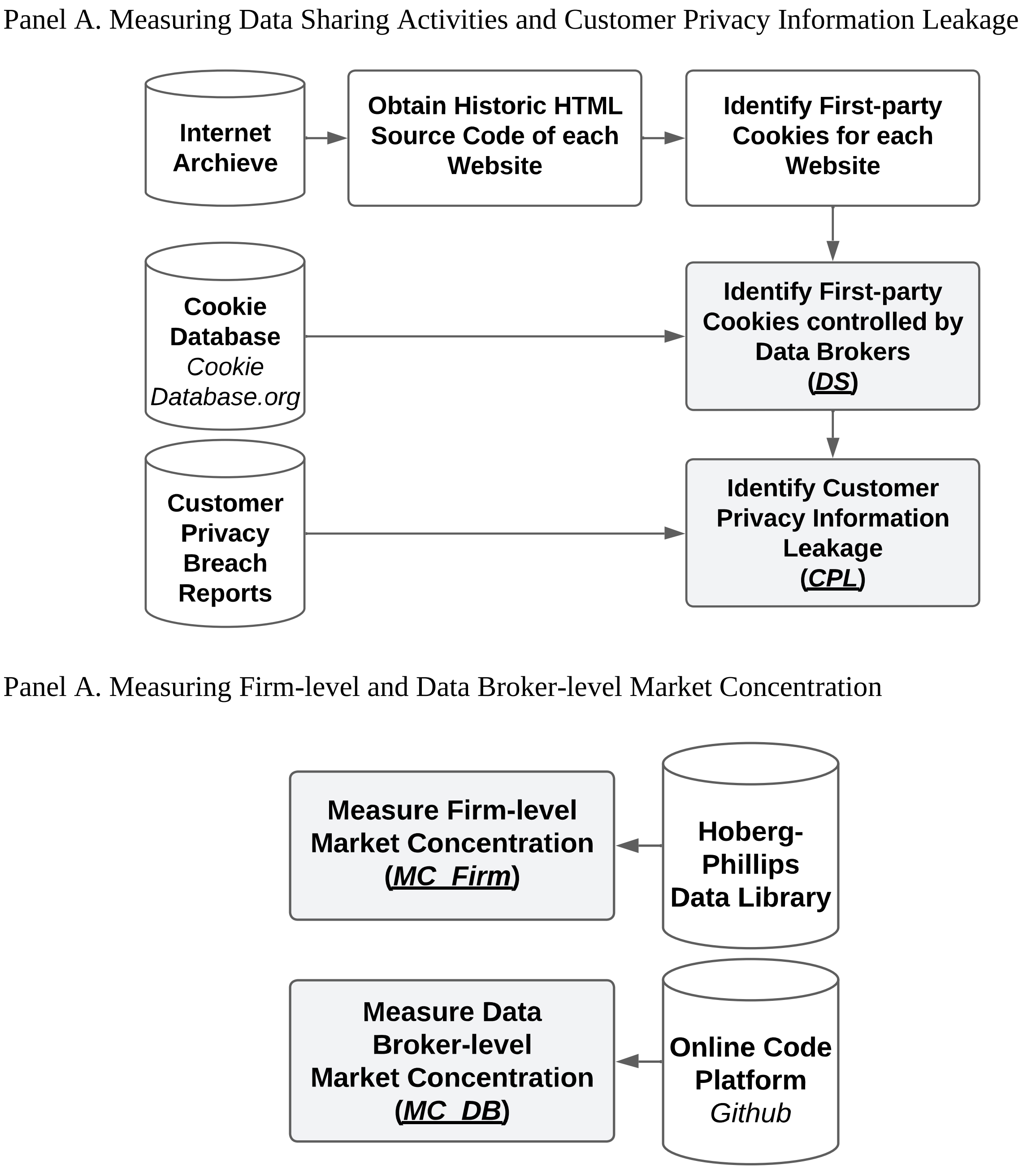


Figure 3. Data Collection and Variable Measurement



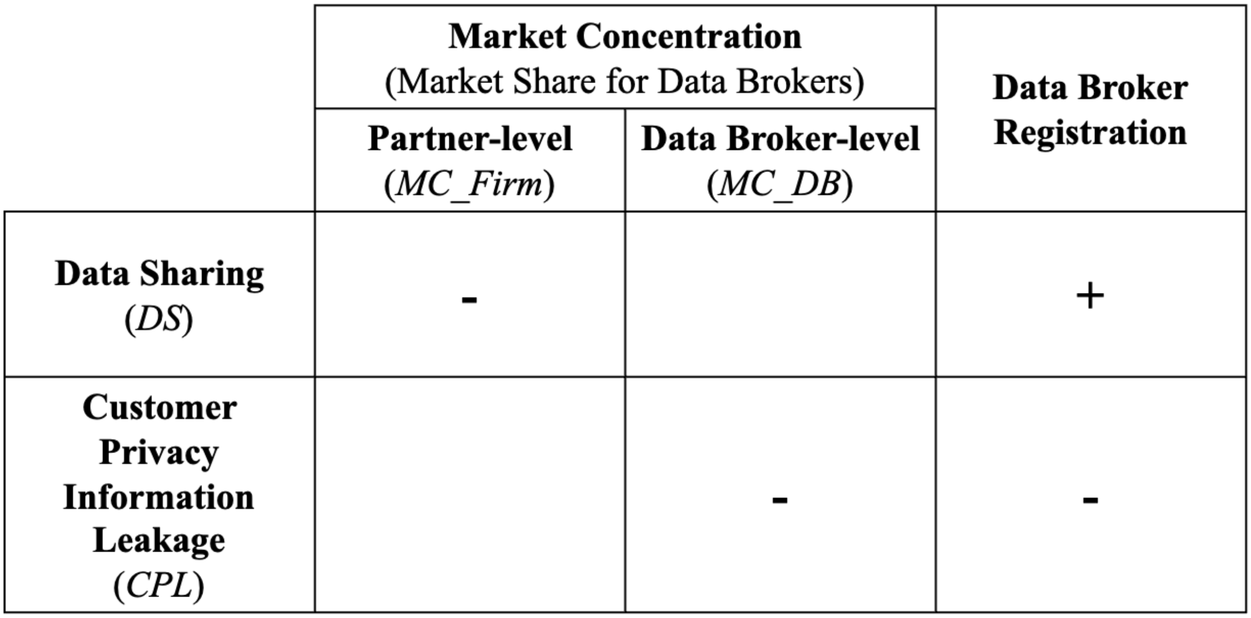
Panel A. Measuring Data Sharing Activities and Customer Privacy Information Leakage

Internet Archieve
Obtain Historic HTML Source Code of each Website
Identify First-party Cookies for each Website
Cookie Database *Cookie Database.org*
Identify First-party Cookies controlled by Data Brokers (*DS*)
Customer Privacy Breach Reports
Identify Customer Privacy Information Leakage (*CPL*)

Panel A. Measuring Firm-level and Data Broker-level Market Concentration

Measure Firm-level Market Concentration (*MC_Firm*)
Hoberg-Phillips Data Library
Measure Data Broker-level Market Concentration (*MC_DB*)
Online Code Platform *Github*

Figure 4. Result Summary



	Market Concentration (Market Share for Data Brokers)		**Data Broker Registration**
	Partner-level (*MC_Firm*)	**Data Broker-level** (*MC_DB*)	
Data Sharing (*DS*)	-		+
Customer Privacy Information Leakage (*CPL*)		-	-

Table 1. First-party Data Collection Activities by Data Brokers[6]

Data Broker	Purpose	Revenue Model
Google AdSense	▪ "Custom Search Ads (including AdSense for Search, AdSense for Shopping, and Programmable Search Engine) also uses a combination of first-party and third-party cookies. First party cookies are relied upon primarily when access to third party cookies is restricted, and are required to continue ad serving."	▪ **Ad Revenue Multiplier** "Custom Search Ads is a Google product that lets you monetize the search results pages of your own search experience. If you don't already have a search experience on your site, consider adding an AdSense search engine, which can provide both a search experience and revenue from search ads."
Oracle	▪ "Online information about you originates from your activities on sites operated by our online partners, such as advertising agencies and website operators … from third parties who may not have a relationship with you and who collect online information using cookies or similar technologies, such as pixels tags."	▪ **Cross-channel Marketing** "Import DMP clients' user attributes into the Oracle Data Cloud platform and help them to leverage and enhance their first-party data for cross-channel marketing.

[6] Google AdSense's revenue model can be found at https://support.google.com/adsense/answer/7549925. Oracle's revenue model is deribed at https://www.oracle.com/legal/privacy/advertising-privacy-policy.html#source.

Table 2. Sample Selection

	Number of Firm-year Observations	*Number of Firms*
Firms with Available Website URL Address	**45,821**	**9,161**
Less: No historical websites available in Archieve.org	(39,578)	(8,034)
Less: Filter broken HTML (smaller than 1,000 words or more than 200,000 words))	(309)	(19)
Less: No financials provided by Compustat	(493)	(81)
Less: No market concentration measure available	(2,918)	(471)
Total Sample (Panel Dataset)	**2,832**	**575**
Less: No unique company or product name to search the darkweb post		**185**
Total Sample (Cross-Sectional Dataset)		**390**

Table 3. Descriptive Statistics

Panel A. Descriptive Statistics

Variable	Count	Mean	Sd/ Mean	Min	P25	P50	P75	Max
Calculated Measures								
CPL	2,523	0.069	4.054	0.000	0.000	0.000	0.000	3.000
MC_Firm	2,523	0.316	0.941	0.020	0.087	0.195	0.452	1.000
MC_DB	2,523	0.517	0.647	0.000	0.231	0.599	0.817	1.000
Control Variables								
NI	2,523	1.174	3.748	-9.261	-2.973	2.137	4.975	9.871
AT	2,523	6.706	0.333	0.646	5.014	6.802	8.398	13.70
Intan	2,523	3.838	0.854	-6.215	0.000	3.885	6.596	12.64
ROA	2,523	-0.091	-4.864	-12.853	-0.089	0.019	0.064	1.285
ROE	2,523	-0.138	-9.009	-51.696	-0.084	0.019	0.056	7.517
Neg	2,523	0.415	1.188	0.000	0.000	0.000	1.000	1.000
Age	2,523	9.481	1.192	0.000	0.000	4.000	21.00	38.00

Notes:

CPL	Firm-level logarithmic value of customer privacy leakage
Cookie	Number of first-party cookies controlled by third-party data brokers.
MC_Firm	Firm-level market concentration (TNIC-HHI)
MC_DB	Data broker level market concentration observed from GitHub
REVT	Firm-level logarithmic value of total revenue
ROA	Firm-level return on assets
ROE	Firm-level return on equity
Neg	Firm-level identifier of negative income (1 if negative income, 0 otherwise)
Age	Firm-level firm age

Table 4. Market Concentration, Data Sharing, and Customer Privacy Information Leakage

Dependent Variable	(1) *DS*	(2) *DS*	(3) *CPL*	(4) *CPL*
Independent Variables				
DS			0.378	1.204**
			(0.451)	(2.010)
MC_Firm	-0.074**	-0.115***	0.0305	
	(-2.045)	(-2.663)	(0.0431)	
MC_DB		-0.022		-0.774
		(-0.964)		(-0.762)
MC_Firm × MC_DB		0.073*		
		(1.664)		
MC_Firm × DS			1.737	
			(1.083)	
MC_DB × DS				-1.129***
				(-2.720)
Control Variables				
NI	-0.045	-0.056	0.102	0.120*
	(-0.130)	(-0.161)	(1.489)	(1.738)
AT	-0.757	-0.73	0.187	0.198
	(-0.866)	(-0.842)	(1.354)	(1.616)
Intan	0.652	0.645	0.171*	0.165**
	(1.156)	(1.143)	(1.868)	(2.051)
ROA	0.929	0.787	-0.271*	-0.226
	(0.708)	(0.589)	(-1.726)	(-1.466)
ROE	0.0633	0.0823	-0.0339	-0.024
	(0.269)	(0.347)	(-0.997)	(-0.782)
Neg	0.398	0.268	0.751	0.995
	(0.138)	(0.092)	(1.147)	(1.499)
Age	0.059	0.0593	0.009	0.012
	(0.339)	(0.335)	(0.645)	(0.793)
Fixed Effects				
Firm / Year / Industry	Included	Included	Included	
Constant	0.114*	0.126*	-7.765***	
	(1.730)	(1.881)	(-7.028)	
Observations	2,523	2,523	2,523	

Dependent Variable	**(1) *DS***	**(2) *DS***	**(3) *CPL***	**(4) *CPL***
Adj R-squared	0.269	0.270		
Wald Chi-squared			41.27	
(Prob>Chi-squared)			(0.000)	

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. *t*-statistics are reported in parentheses.

Table 5. The Effectiveness of Data Broker Registration

Dependent Variable	**(1) *DS***	**(2) *CPL***
Independent Variables		
DS		0.292
		(0.281)
MC_Firm	-0.060**	
	(-2.251)	
MC_DB		-1.264***
		(-2.939)
DB_Reg	0.221**	-0.090
	(2.128)	(-0.286)
DB_Reg × DS		0.234
		(1.630)
DB_Reg × MC_Firm	0.0580	
	(0.385)	
DB_Reg × MC_DB		4.171***
		(7.844)
MC_DB × DS		0.965
		(0.634)
DB_Reg × MC_DB × DS		-4.257***
		(-10.21)
Control Variables		
NI	-0.006	0.122*
	(-0.031)	(1.737)
AT	-0.078	0.216*
	(-0.142)	(1.716)
Intan	0.0163	0.162**
	(0.044)	(1.999)
ROA	-0.164	-0.235
	(-0.221)	(-1.503)
ROE	0.128	-0.0232

Dependent Variable	**(1)** ***DS***	**(2)** ***CPL***
	(0.945)	(-0.748)
Neg	0.938	1.024
	(0.457)	(1.505)
Age	0.007	0.0156
	(0.071)	(0.984)
Fixed Effects		
Firm / Year / Industry	Included	Included
Constant	0.0516	-7.476***
	(1.072)	(-7.335)
Observations	2,523	2,523
Adj R-squared	0.650	
Wald Chi-squared		196.29
(Prob>Chi-squared)		(0.000)

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. *t*-statistics are reported in parentheses.

Table 6. Clustering by Data Sharing Activities

Panel A. K-Means Clustering Within-Cluster Sum of Square (WCSS)

Number of Clusters	2	3	4	5	6	7	8	9	10
WCSS	89.45	56.42	38.55	27.55	17.87	11.60	4.93	2.93	0.99

Panel B. Cluster Results

Cluster	*N*
1	50
2	110
3	27
4	22
Total	209

Table 7. ANOVA Result of Data Sharing Activities

Panel A. ANOVA Result

Dependent Variable = *DS* (*N*=209)	***Firm-level*** (*Adj R*^{2}: 0.469)			***Data Broker-level*** (*Adj R*^{2}: 0.444)		
	Partial SS	*df*	*MS (F)*	*Partial SS*	*df*	*MS (F)*
Model	20.70	7	2.958^{***} (27.33)	19.66	7	2.809^{***} (24.76)
Market Concentration	0.586	1	0.586^{**} (5.410)	0.061	1	0.614^{**} (0.540)
Cluster	17.20	3	5.735^{***} (52.98)	19.42	3	6.476^{***} (57.09)
Market Concentration × Cluster	1.180	3	0.393^{**} (3.630)	0.180	3	0.060 (0.530)
Residual	21.76	201	0.108	22.80	201	0.113

Panel B. Firm-level ANOVA-adjusted Means (Delta-method)

Dependent Variable = *DS* (*N*=209)		***Margin (Std. Err)***
Market Concentration (Firm)	*High*	1.226^{***} (0.040)
	Low	1.375^{***} (0.049)
Cluster	*1*	1.000^{***} (0.054)
	2	1.008^{***} (0.031)
	3	1.000^{***} (0.063)
	4	2.194^{***} (0.090)

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. t-statistics are reported in parentheses.

Table 8. ANOVA Result of Customer Privacy Information Leakage

Panel A. ANOVA Result

Dependent Variable = *CPL* (*N*=209)	***Firm-level***(*Adj R*^{2}: 0.053)			***Data Broker-level***(*Adj R*^{2}: 0.061)		
	Partial SS	*df*	*MS (F)*	*Partial SS*	*df*	*MS (F)*
Model	0.815	7	0.116** (2.690)	0.886	7	0.126*** (2.950)
Market Concentration	0.096	1	0.096 (2.240)	0.328	1	0.328*** (7.650)
Cluster	0.256	3	0.085 (1.980)	0.239	3	0.079 (1.860)
Market Concentration× *Cluster*	0.409	3	0.136 (3.150)	0.455	3	0.151** (3.530)
Residual	8.705	201	0.0433	8.634	201	0.042

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. t-statistics are reported in parentheses.

Panel B. Firm-level ANOVA-adjusted Means (Delta-method)

Dependent Variable = *Breach* (*N*=209)		***Margin (Std. Err)***
Market Concentration (Data Broker)	*High*	0.011 (0.024)
	Low	0.107*** (0.024)

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. t-statistics are reported in parentheses.

Table 9. Market Concentration, Data Sharing, and Customer Privacy Information Leakage (Observed from the Dark Web)

Dependent Variable	**(1) *CPL_DW***	**(2) *CPL_DW***
Independent Variables		
DS	0.074	0.093^{**}
	(0.37)	(0.25)
MC_Firm	0.071	
	(0.98)	
MC_DB		-1.906
		(1.22)
MC_Firm × *MC_DB*		
MC_Firm × *DS*	-0.139	
	(-1.12)	
MC_DB × *DS*	-0.139^{***}	-0.833^{***}
	(-1.12)	(-2.61)
Control Variables		
NI	0.002	-0.004
	(0.24)	(-0.48)
AT	0.022	0.019
	(1.54)	(1.36)
Intan	0.010	0.006
	(1.09)	(0.72)
ROA	-0.069	-0.075
	(-0.89)	(-1.00)
ROE	0.008	0.021
	(0.34)	(0.84)
Neg	0.039	-0.023
	(0.46)	(-0.26)
Age	-0.001	-0.000
	(-1.02)	(0.09)
Fixed Effects		
Firm / Year / Industry	Included	Industry

Dependent Variable	**(1) *CPL_DW***	**(2) *CPL_DW***
Constant	-0.053	-1.906
	(-0.49)	(-1.56)
Observations	2,523	390
Adj R-squared	0.047	0.070
Wald Chi-squared	32.29	
(Prob>Chi-squared)	(0.000)	

Note: ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. *t*-statistics are reported in parentheses.