12-2021

The Invisible Risk: The Data-sharing Activities of Data Brokers and Information Leakage

Arion Cheong

Tawei (David) Wang

The Invisible Risk: The Data-sharing Activities of Data Brokers and Information Leakage

Completed Research Paper (Extended Abstract)

Arion Cheong
California State University at Fullerton
acheong@fullerton.edu

Tawei Wang
DePaul University
david.wang@depaul.edu

Abstract

Data brokers are the major players in the market for collecting, selling, and sharing user information. This paper considers data brokers' data sharing activities as a co-opetition between data brokers and investigates how the information collecting and sharing activities may lead to information leakage on the dark web. We find that S&P 1,500 firms experience higher information leakage when sharing more customer information with data brokers through third-party cookies. Further, using the registered data brokers and their competitors as the sample, we observe that registered data brokers are more susceptible to information leakage with data sharing activities than unregistered data brokers. Our study provides initial evidence on the consequences of data brokers' data sharing activities.

Keywords

Data brokers, data sharing, personal information leakage, co-opetition.

Introduction

Organizations have been collecting and sharing an unprecedented amount of information in recent years. Such data collection and sharing activities have raised privacy concerns regarding what information has been collected and how those pieces of information have been used (Mims, 2018). While the public's focus stays with the organizations that have direct relationships with the consumers (e.g., Facebook), the revelation of the existence of data brokers surprises the regulators and the public. The U.S. Fair Trade Commission (FTC) defines data brokers that "typically collect, maintain, manipulate, and share a wide variety of information about consumers without interacting directly with them" (FTC, 2014). That is, these organizations, though they stay behind the scenes, play a critical role in the information market by collecting and sharing/selling user information with other organizations or other data brokers.

In response to this murky situation, regulators have taken action to require the data brokers to disclose more information about their practices to the public, inform the types of information collected, and guide the process to opt-out from their data collection and sharing activities. Two notable examples are Vermont's Data Broker Law and California's Data Broker Law, by which data brokers are now required to register with the Secretary of State in Vermont or Attorney General in California.

In our paper, we investigate 1) whether the data sharing activities of S&P 1,500 firms with data brokers associate with personal information leakage and 2) the effectiveness of data broker registration to deter personal information leakage when data brokers share information with other data brokers. In specific, we first examine whether the amount of data exchange activities of S&P 1500 firms with data brokers positively associates with information leakage. We measure the data exchange activities of S&P 1,500 firms with data brokers by identifying the third-party cookies in each firm's website, which is known to collect and share customer information with data brokers. Second, we analyze whether the registered data brokers engage more in data sharing activities and see how such activities lead to personal information leakage.

Our findings suggest that firms with higher data exchange activities with data brokers have more information leakage. Further, we find that registered data brokers, even providing easy access to opt-out for information collection, are more susceptible to information leakage when they engage more in data sharing activities.

In the remaining sections of this study, we provide a summary of relevant literature and our hypotheses. Following that, our analyses and the results are presented. We conclude by providing policy implications based on our results.

Data Sharing and Co-opetition

Data brokers provide their collected information to the downstream client firms or sell it back to the first-party data holders, who use it in a number of ways. Prior studies focused on the strategic behaviors of data brokers for sharing or selling their information to other parties. Accordingly, Gu et al. (2019) define the data broker market in a context as "co-opetition," where the data brokers compete to supply data to client firms in the market while cooperate and share data with other competing data brokers to maximize their profit.

In our first hypothesis, we aim to provide direct insight into the third-party risk posed to the firm (i.e., S&P 1,500 firms) when they share their customer's information with the data brokers. Our major concern about these firms is their data exchange activities with data brokers, which possess a high risk of information leakage, without any notifications given to the customers about it. We hypothesize this concern to examine how data sharing activities between S&P 1,500 firms and data brokers lead to personal information leakage. Formally, we state our first hypothesis:

H1: The amount of data exchange activities of firms with data brokers is positively associated with information leakage.

Data Broker Registration and the Pre-collection Notice Exemption

Under California's data broker law, only the registered data brokers are allowed to sell or share onward their collected personal information without a pre-collection notice (Eisert 2020). However, many data brokers still go unregistered in Vermont and California data broker registries due to the lack of enforceability (Ruhaak 2019). In fact, there is no case of prosecution taken by either Vermont or California Attorney General regarding registration. The penalty is extremely low for unregistered firms, e.g., unregistered firms in California may be subject to a penalty of $100 per day. Accordingly, experts address that the loosely defined data broker allows the data brokers to selectively register as a data broker (Sherman 2021).

Given that registration is not mandatory in practice, we believe the registration is a strategic means to obtain the pre-collection notice exemption that allows the registered data brokers to collect and share information in the dark. While expecting more data sharing activities of registered data brokers, it is intuitive that a firm's information will have a higher chance to be leaked when it is shared with registered data brokers. Therefore, registered data brokers have higher information leakage since the attack surface becomes enlarged by data exchange between data brokers. In accordance, we extend our first hypothesis to examine whether registered data brokers are susceptible to information leakage due to their data sharing activities.

H2: Registered data brokers with more data sharing activities have higher information leakage than unregistered data brokers.

Measuring Data Sharing Activities and Information Leakage

Based on our research questions, we have collected three sets of data: the data-sharing activity of firms (S&P 1,500), the data-sharing activities between data brokers and their coopetitors, and the information leakage history of S&P 1,500 firms and data brokers. To measure the data sharing activities of S&P 1500 firms, we scrape the cookies from each firm's website. In specific, we use *Selenium* to collect the cookies that are an application commonly used to test and verify the elements and function of a website. Our sample includes only firms that have a website that directly interacts with customers (i.e., requiring the customer to log in for identification). Based on the cookies that are collected, we identify third-party cookies that

advertisement firms and data brokers widely use to collect customer information. Further, we identify information leakage from the darknet market posts that were uploaded between a six-month period, from the next day of the annual data broker registration deadline (February 1st, 2020) to July 31st, 2020. In our sample, we selected 296,512 posts that contain at least one personally identifiable information (PII), which was detected by the algorithm (i.e., classified to include hacked information).

Results

To examine our first hypothesis, we construct a regression model to study the effect of data sharing activities on information leakage. In specific, we first measure the level of data sharing activities for each non-data broker firm by counting the number of third-party cookies. Next, we count the number of dark web posts containing the firm's unique identifier within a six-month period. In addition to the baseline model, we consider control variables relates to the firm's internal control. Our result of the regression analysis shows that more amount of third-party cookie collection leads to higher customer information conveyed to data brokers. In specific, S&P firms with more data sharing activities with data brokers show higher information leakage that supports our first hypothesis (H1).

We further examine the effect of data sharing activities between data brokers on information leakage while considering the registration status. To correspond the systematic difference in information leakage with registration status, we perform a regression analysis with an interaction term between the number of coopetitors, which represents the degree of data sharing activities between the data brokers, and the registration status. The results shows that registered data brokers are more susceptible to information leakage when more data sharing activities occur, compared to the unregistered data brokers.

Conclusion

The business of collecting and sharing data by data brokers is not regulated, and the risk is not well known and informed to the public. This study provides empirical evidence about the positive association between information leakage and data sharing activities between firms and data brokers. We also show that registered data brokers are more susceptible to information leakage when they engage more in data-sharing activities. Based on our result on systemic risk shared among data brokers, the regulators should introduce an effective framework to ask the data brokers to enhance their disclosure about their data collection activities and the risk associated with their data-sharing activities.

References

Eisert, R. 2020. "The California Data Broker Registry's Growing Significance For Ad Tech," AdExchanger (https://www.adexchanger.com/data-driven-thinking/the-california-data-broker-registrys-growing-significance-for-ad-tech/; accessed Nov 2021).

Federal Trade Commission (FTC). 2014. "Data brokers: A call for transparency and accountability," Federal Trade Commission Washington (https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf; accessed Nov 2021).

Gu, Y., Madio, L., and Reggiani, C. 2019. "Data brokers co-opetition," *CESifo Working Paper (7523)*, Center for Economic Studies and ifo Institute (CESifo), Munich.

Mims, C. 2018. "Who has more of your personal data than Facebook? Try Google," The Wall Street Journal (https://www.wsj.com/articles/who-has-more-of-your-personal-data-than-facebook-try-google-1524398401; accessed Nov 2021).

Ruhaak, A. 2019. "Data Brokers are Cruising for a Bruising," Wired (https://www.wired.com/story/opinion-data-brokers-are-cruising-for-a-bruising/; accessed Nov 2021).

Sherman, J. 2021. "Federal Privacy Rules Must Get "Data Broker" Definitions Right," Lawfare (https://www.lawfareblog.com/federal-privacy-rules--must-get-data-broker-definitions-right; accessed Nov 2021).