How Much to Share with Third-Parties?
A Website's Dilemma and Users' Privacy Concerns

Ram Gopal[1], Hooman Hidaji[2], Raymond A. Patterson[3], Erik Rolland[4], and Dmitry Zhdanov[5]

[1] University of Connecticut; ram.gopal@business.uconn.edu
[2] University of Alberta; hooman.hidaji@ualberta.ca
[3] University of Alberta; ray.patterson@ualberta.ca
[4] University of California, Merced; erolland@ucmerced.edu
[5] University of Connecticut; dmitry.zhdanov@business.uconn.edu

Abstract

Websites are increasingly presenting content and services that are not created by the publisher website administrators, but are provided by other "third-parties" and have different purposes. While these components provide potential utility to the visitor, they come at the cost of visitor information being shared. With growing concerns regarding online privacy and information disclosure, it is important to know what factors affect these connections between website publishers and third-parties. The privacy concerns surrounding the leakage of information have been growing rapidly. In this study, we propose a two-sided economic model that captures the interaction between the users, publisher websites, and third-parties. Specifically, we focus on the effect of privacy concerns on the sharing behavior of the website. We also analyze the welfare aspects of such concerns. Finally, we provide insight on industry regulations and policy. We validate the model using an exploratory empirical analysis of sharing with third-parties and industry structure among these players. To the best of our knowledge, this study is among the first to analyze the decision-making in sharing of user information on websites.

1. Introduction

Sites on the World Wide Web have become ever more resourceful, enabling visitors to obtain various services and information from them. In doing so, publisher websites might outsource, and present content and services provided by third-party providers on their pages. Thus the experience of visiting a website by a user involves interactions with many third-parties. Third-party components within publisher websites enable content and services such as targeting and advertising (e.g., ad presentation and analytics), functionality (e.g., password security, social media integration, video hosting, chat and forum services, and payment services), and site performance improvements (e.g., backup service, website security, and responsiveness tools), and thus add to the usability of these websites. Figure 1 depicts a publisher website (engadget.com) highlighting its various visible third-party components.

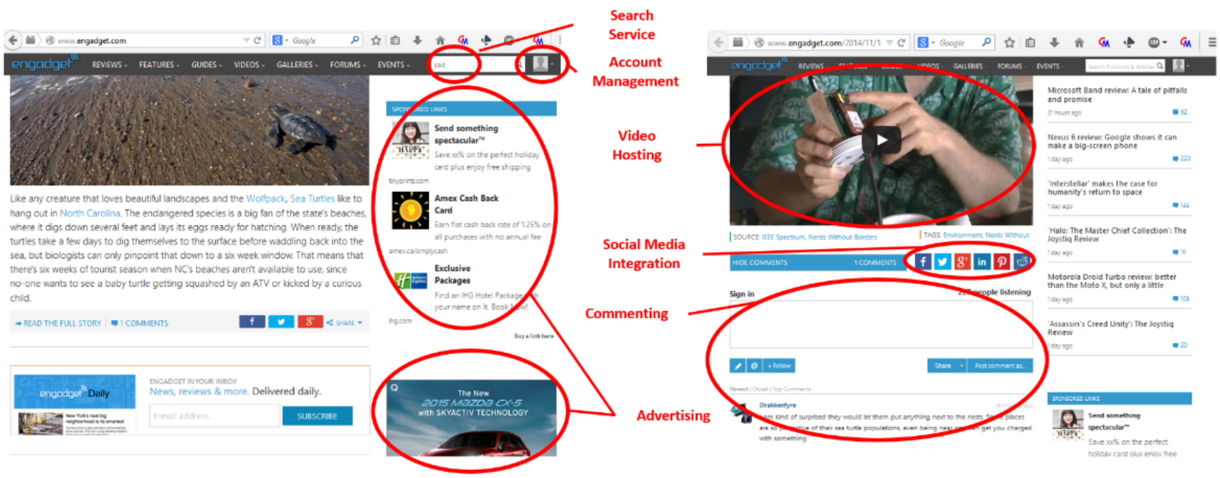

Figure 1. A Sample Publisher Website with Third-Party Content Highlighted

Third-parties can and do collect visitor information from publisher websites, and there is a potential tradeoff between the use of third-party components and visitor privacy. In many cases, cookies are used by the third-parties to track user activity. The rise in use of third-party content is in tandem with the growing number of people going online (Smith 2014). It has become extremely hard, if not impossible, to know who is tracking visitors online (Schoen 2009). It is also noted that the extent of collected information is not limited to browsing data, and in fact the information can be used for identification or re-identification of individuals when used alongside other sources (Krishnamurthy and Wills 2009a). Privacy issues that arise from the increased usage of third-parties and cookies is a public concern, and is being investigated by authorities and policy makers such as the Federal Trade Commission and European Union (Mayer and Mitchell 2012).

While the problem of information privacy in websites is faced by many, it has not received much attention from researchers. Many have studied information privacy in the context of e-commerce where users willingly provide their information to companies (Li 2012), but there is a gap in the literature concerning the use of third-parties by publisher websites and the involuntary disclosure of personal information. In this paper, we fill the gap by considering the issue of monetization versus information privacy from an economic viewpoint. The contributions of this paper are as follows. We provide a two-sided economic model that describes the decision-making process of the publisher website based on the privacy concerns of the user, participation incentives of third-party service providers, and the publisher website's own incentives to maximize profits. We discuss the effects of privacy concern on the stakeholders and implications for policy-making. Finally, we provide empirical validation and support for several important aspects of the model. This empirical validation also serves to illustrate the problems surrounding information privacy versus website monetization that Internet publisher websites, users, third-parties, and policy-making organizations are collectively facing. The proposed model successfully explains differences in third-party sharing by websites in contexts with different needs for privacy, and provides managerial and policy insights for publisher websites, policy-making organizations, and governments.

The paper is organized as follows. Literature review is provided in section 2. The model is presented in section 3. Section 4 presents an analysis of market structure and implications for public policy. Section 5 presents a validation of key model results using an exploratory empirical study of third-party sharing by publisher websites, and section 6 concludes the paper.

2. Literature Review

Despite the omnipresent use of third-parties in publisher websites, there are not many papers that study the effects and issues around this matter. Third-parties do provide some benefit by providing additional services to websites which can be inferred by the pandemic use of third-parties in websites (Mayer and Mitchell 2012). Chen and Stallaert (2014) provide an economic analysis of online behavioral advertising, which is the most popular type of third-party use. They find the conditions in which the use of behavioral targeting in advertising using user information can be beneficial to the publisher website. While their model can incorporate the privacy concerns of users in the form of opting out of the service, the authors do not consider the problem from a privacy point of view.

There are many papers that study the effect of third-parties on user information diffusion. Krishnamurthy and Wills (2006) study the diffusion of privacy on the Internet, looking at the possibility of aggregating user information from different sources. They study popular websites and conclude that "the size of the privacy footprint is a legitimate cause for concern" across the sites studied. They also find significant increase in privacy footprint over a six month period. Krishnamurthy and Wills (2009b) show in a longitudinal study that the sharing and aggregation of user information has been increasing, while the number of entities involved has been decreasing as a result of acquisitions.

One of the issues that arise as a result of privacy leakage, is discrimination among different users. Krishnamurthy et al. (2011) study websites that require users to register and provide information, and find that 75% of the popular websites studied leak sensitive user information to third-parties. They conclude that "the growing disconnect between the protection measures and increasing leakage and linkage suggests that we need to move beyond the losing battle with aggregators and examine what roles first-party sites can play in protecting the privacy of their users." Mikians et al. (2012) and Valentino-DeVries et al. (2012) among others, provide evidence for price and search discrimination in e-commerce setting, which is based on user information on the web. Recently, there has been even more concern about the implications of such information sharing. A recent analysis found that even a government health care website is sharing important user health and other information with third-parties (Alonso-Zaldivar and Gillum 2015).

There are many privacy blocking tools available for reduction or prevention of third-party sharing. Krishnamurthy et al. (2007) study the effect of blocking tools in protecting privacy when browsing the Internet, and the effect on website quality and usability. They find that while first-party privacy blocking has some effect on the usability of websites, third-party blocking has minimal effect on website quality. Leon et al. (2012) study several tools for limiting online behavioral targeting, the most popular use of third-parties, and found serious flaws with them that results in users' inability to control their privacy settings. Malandrino and Scarano (2013) study how third party sites collect and aggregate data and build personal profiles of unaware users. They provide an empirical study on how user's privacy can be undermined because of such involuntary privacy violations, and experiment with tools that can inform users and give them control over such activities. However we should note that these tools are usually used by enthusiastic users, and not by the majority of users.

The literature on user preferences shows that they are concerned about the privacy violations of third-party information sharing. Turow et al. (2009) surveys users in the United States to find that 68-87 percent of them do not want to be tracked for advertising purposes. McDonald and Cranor (2010) also find that only 20% of users prefer targeted online advertising over random advertising. Mayer and Mitchell (2012) provide a review of the policies and technology surrounding the third-party information sharing, and note that it is crucial for researchers to do perform more research on the issues.

Online information privacy has been studied by many researchers. Li (2012) provides a comprehensive review of the extensive online information privacy literature, and provides a framework for theoretical research on the user's privacy decision-making. While this literature does not directly study third-parties, many of the principles are applicable to information privacy in third-party sharing. User information disclosure in the form of third-parties usage can be explained by theories such as privacy calculus theory, risk calculus theory, and dual-calculus theory among many foundational theories (Li 2012). In this paper, we are not concerned with the user's information disclosure process, and presume that users are able to calculate their privacy costs and include them in their utility function.

On the other hand, the publisher website behaviour cannot be viewed in isolation, as it is affected by both users and third-parties. The publisher realizes that the user can estimate its privacy cost, by using a model such as dual privacy calculus theory (Li 2012), and it makes decisions in response to the user's perception of benefits and privacy risks. According to agency theory and utility maximization theory and their application in information privacy (Li 2012), a publisher website would sets its decision variables so as to maximize total profit from users and third-parties collectively.

In terms of modeling, this paper is related to the literature in two-sided markets, where a platform provider is affected by two markets which have some kind of interaction or network externality. Most of the studies consider markets in which positive indirect network effects are present among the two markets. Parker and Van Alstyne (2005) and Rochet and Tirole (2003) study the pricing strategies in such markets. Anderson et al. (2013) consider the platform investment in two-sided networks. We use a similar model setting to Anderson et al. (2013). However, in this paper, we consider a two-sided market where both positive and negative network effect are present among the two markets.

While some of the mentioned studies look at the information flow among websites and third-parties, they do not provide much insight on the decision-making process of the websites in doing so, and how the privacy concerns affects those decisions. We next provide a two-sided economic model that explains the decision-making of publisher websites with respect to using third-parties. We especially study how privacy concerns affect the publisher website, users, and third-parties. To the best of our knowledge, this is the first attempt to model the publisher website's decision making.

3. Model

In this section, we propose an economic model in order to describe and analyze the problem of third-party usage in websites. The notation is provided in Table 1 and Table 2.

3.1 Model Setting

Here we use a two-sided model that involves the three players. The publisher website is seen as a platform where users and third-parties interact. The analysis is provided for a single publisher website, and for one or more third-parties and users. The website is seen as a platform in this sense. So the model is a price-maker monopoly with violation and subscription price as decision variables. The model incorporates two-sided network effects of users and third-parties. Anderson et al. (2013) use a similar setting.

In our model, the third-party benefits from the users on the website. It is assumed that the more users there are, the more information the third-party can collect, which is beneficial. On the other hand, we assume that third-parties provide no additional benefit to the user. While third-parties do provide some utility to the user, here we consider the case in which the publisher website is considering whether or not to outsource a service on the website, where the third-party is a substitute for the website's own service. So the assumption is that the third-party brings only violation to the user with no additional benefit to the user than what they already receive from the website in terms of intrinsic value x. As discussed in section 2, there are many studies that provide evidence for the negative utility of third-parties for users (Krishnamurthy and Wills 2006, Turow et al. 2009). Moreover, Krishnamurthy et al. (2007) found that blocking of third-parties does not significantly affect the usability of websites. The utility function of the user is given as:

$$U(x) = x - N_D \, v \, V - P_W \tag{1}$$

where V is the amount of privacy violation inflicted on the user from each unit of service as decided by the website (privacy cost to user as a result of using third-party content). We use the term "violation" to allude to all of the possible ways that data can be leaked or linked, and that detracts from the user's utility if is understood by the user. Violation is defined by both the actual violation and the user's perception of that violation. v is user's disutility from an additional unit of privacy violation from each third-party, or in other words, user's sensitivity to privacy violations. It can also be seen as the violation that is perceived by the user, or is visible to the user. P_W is subscription fee of the publisher website as set by the website. Both V and P_W are the decision variables of the publisher website.

Table 1. Model Parameters and Variables

Notation	Definition
$U(x)$	Utility of a user with intrinsic value of x from the website, a user will use the website only if $U(x) \geq 0$.
Π_D	Third-party's profit, $\Pi_D \geq 0$
Π_W	Website's profit, $\Pi_W \geq 0$
x	Intrinsic value of the website for a user $x \geq 0$
X	Maximum intrinsic value of the website for user, $X > 0$
f	Fixed cost of a third-party, $f \geq 0$
F	Maximum third-party fixed cost, $F > 0$
N_D	Number of third-parties on the website, $N_D \geq 0$
M_D	Total number of third-parties in the market, $M_D \geq N_D, M_D > 0$
N_U	Number of users who use the website, $N_U \geq 0$
M_U	Total number of users in the market, $M_U \geq N_U$, $M_U > 0$
v	User's perceived disutility from an additional unit of privacy violation from each third-party, $v \geq 0$. This is the privacy concern of the user, which emanates from the portion of the violation that is visible to the user.
R_V	Third-party's net revenue from each unit of violation from each user, $R_V \geq 0$
C_W	Publisher website's per user operating costs, $C_W \geq 0$
F_W	Publisher website's fixed operating costs, $F_W \geq 0$
r	Royalties paid to the publisher website by third-party, $r \geq 0$

Table 2. Model Decision Variables

Notation	Definition
V	The amount of privacy violation to user from each third-party $0 \leq V < \infty$
P_W	Price of using the publisher website for each user, $0 \leq P_W < \infty$

Let $\hat{x}$ be the user who is indifferent between using and not using the website. For the indifferent user we have $\hat{x} = P_W + N_D v\ V$ and if we assume that x is uniformly distributed over $[0, X]$, then the participation rate of the users is calculated as $(X - \hat{x})/X$. We assume that there are M_U users in the market, that is the total number of potential users. The number of users that will use the publisher website, N_U is given by:

$$N_U = \frac{M_U (X - N_D\ v\ V - P_W)}{X} \geq 0 \tag{2}$$

We also assume that the third-party pays royalties to the publisher website that is independent of the number of users on the website. So the profit function of the third-party is as follows:

$$\Pi_D = N_U\ V\ R_V - r - f \tag{3}$$

where R_V is the net revenue from each unit of user information, r is the one-time royalty paid to the publisher website by third-party, and f is the fixed cost of the publisher, assumed to be uniformly distributed over $[0, F]$. The marginal developer $\widehat{f}$ who is indifferent between entering or not entering the publisher website is characterized by $\widehat{f} = N_G\ V\ R_V - r$. The third parties having $f < \widehat{f}$ will join the website, so the number of third-parties that will provide the service is calculated as $\widehat{f}/F$. Assuming that there are overall M_D number of third-parties in the market, the number of third-parties that will join the market, N_D is calculated as:

$$N_D = \frac{M_D (N_U\ V\ R_V - r)}{F} \geq 0 \tag{4}$$

It can be seen that there are two-sided network externalities present among the number of users and number of third-parties, in a way that third-parties prefer more number of users, while users prefer less number of third-parties, and the monopolist publisher website benefits from both. Solving formulae for N_U and N_D we have:

$$N_U = \frac{M_U (M_D\ r\ v\ V + F(X - P_W))}{M_D\ M_U\ R_V\ v\ V^2 + FX} \geq 0 \tag{5}$$

$$N_D = \frac{M_D (M_U\ R_V\ V(X - P_W) - r\ X)}{M_D\ M_U\ R_V\ v\ V^2 + FX} \geq 0 \tag{6}$$

Note that the number of third-parties and number of users are defined in terms of participation ratio. We use the terms number and participation ratio interchangeably in this context.

It is assumed that the amount of violation, V, is decided by the publisher website. This can be true in the case where there are numerous third-parties with different known violation factors available in the market to choose from, so the publisher website can combine third-parties to achieve the required violation factor. The profit function of the publisher website in this case is as follows:

$$\Pi_W = N_U\ (P_W - C_W) + N_D\ r - F_W \tag{7}$$

By substituting formulae for N_U and N_D from (5) and (6) we have:

$$\Pi_W = \frac{M_U(P_W - C_W)(M_D r\, v\, V + F(X - P_W)}{M_D M_U R_V v V^2 + FX} + \frac{M_D r\,(M_U R_V V(X - P_W) - r\,X)}{M_D M_U R_V v V^2 + FX} - F_W \tag{8}$$

By solving the first and second order conditions we can find the optimal violation and publisher website price as provided in Lemma 1 below. The values for violation and publisher website price are assumed to be non-negative.

Lemma 1.

The platform monopolist chooses

$$V^* = \frac{r\,X\,(R_V + v)}{M_U R_v\, v\,(X - C_W)} \geq 0$$

$$P_W^* = \frac{F\,M_U R_v\, v\,(X^2 - C_W^2) - X\,M_D r^2 (R_V^2 - v^2)}{2\,F\,M_U R_v\, v\,(X - C_W)} \geq 0$$

◊

Proof for Lemma 1 is provided in Appendix 1. Next, we discuss the interaction among different variables of the model.

3.2 Effect of Model Parameters on Publisher Website Decision-Making

Using the Lemma 1, Propositions 1 and 2 below provide the effect of different parameters on the publisher website decision variables of optimal violation factor V^* and website price P_W^*. As provided in proof for Lemma 1 and in propositions 2 and 3, we assume the user participation constraint of $X - C_W > 0$ and the third-party participation constraint of $R_V - v \geq 0$.

Proposition 1.

The optimal violation factor V^* and the optimal website price P_W^* satisfy the following:

(i) V^* decreases with users' perceived privacy violation from each third-party (v) at an increasing rate.

(ii) V^* increases with publisher website royalties paid by third-party (r).

(iii) V^* decreases with third-party revenue from each unit of violation from each user (R_V) at an increasing rate.

(iv) V^* increases with website operating costs per user (C_W) at an increasing rate.

(v) P_W^* increases with users' perceived privacy violation from each third-party (v) at a decreasing rate.

(vi) P_W^* decreases with publisher website royalties paid by third-party (r) at a decreasing rate.

(vii) P_W^* decreases with third-party's revenue from each unit of violation from each user (R_V) at an increasing rate.

(viii) P_W^* *increases with website operating costs per user* (C_W) *when* $F\ M_U\, R_V\ v\ (X-C_W)^2 - X\, M_D\, r^2\, (R_V^2 - v^2) > 0$, *and decreases with it when* $F\ M_U\, R_V\ v\ (X-C_W)^2 - X\, M_D\, r^2\, (R_V^2 - v^2) < 0$*, at a decreasing rate.*

◊

Proposition 1 part (i) is shown to be true by the first derivative:

$$\frac{\partial V^*}{\partial v} = -\frac{r\,X}{M_U v^2\,(X-C_W)}$$

which is negative when $X - C_W > 0$. This intuitively provides that when users' perception of the violations (v) are higher, publisher website sets a lower violation factor (V^*) so as to maintain a desired level of perception combined with real violation (vV^*). In terms of privacy, this states that for websites whose users are more concerned about their privacy, or where sensitive information is present, the website reduces the amount of privacy violation. By taking the second derivative we have:

$$\frac{\partial^2 V^*}{\partial v^2} = \frac{2\,r\,X}{M_U v^3\,(X-C_W)}$$

which is positive when $X - C_W > 0$. This means that publisher website's response to decrease the actual violation factor (V^*) when user perception of privacy violations (v) increases, is increasingly larger for larger changes in v. Interpreting this, same incremental changes in user perception of privacy violation (v) have increasingly greater impact on the actual privacy violation (V^*).

For part (ii) of the proposition, the first derivative is as follows:

$$\frac{\partial V^*}{\partial r} = \frac{X\,(R_V+v)}{M_U\, R_v\, v\,(X-C_W)}$$

which is positive when the participation condition of $X - C_W > 0$ binds. This proposition depicts an interaction between the publisher website and the third-parties, whereby in exchange for increased royalties, the website needs to let third-parties violate user privacy. The second derivative is zero, so this occurs at a constant rate.

Part (iii) of the proposition is shown by the first derivative:

$$\frac{\partial V^*}{\partial R_V} = -\frac{r\,X}{M_U\, R_v{}^2\,(X-C_W)}$$

which is negative when participation condition of $X - C_W > 0$ binds. This shows that when third-parties can make higher per unit revenue from information violation, it is beneficial for the publisher website to reduce the amount of actual violation. Taking the second derivative of V^* with respect to R_V we have:

$$\frac{\partial^2 V^*}{\partial R_V{}^2} = \frac{2\,r\,X}{M_U\, R_v{}^3\,(X-C_W)}$$

which is positive when $X - C_W > 0$. As the information violation revenue (R_V) gets larger, the publisher website's response to decrease the actual violation factor (V^*) becomes even larger. This illustrates a natural braking system to the actual violation. Intuitively, it seems logical that the relationship would be in the opposite direction. We might expect that higher per unit revenue earned by third-parties would actually encourage more violation. However, the counterintuitive finding is that higher per unit violation revenues for third-parties will have a braking effect on actual violation by reducing the violation at an increasing rate. Note also that the perceived violation is not involved in this relationship.

Part (iv) is shown by taking the first derivative:

$$\frac{\partial V^*}{\partial C_W} = \frac{r\,X\,(R_V+v)}{M_U\,R_v\,v\ \ (X-C_W)^2}$$

The first derivative is always positive. When the operational cost of publisher website (C_W) is high, the website is encouraged to violation user privacy in order to remain profitable. The second derivative provides that:

$$\frac{\partial^2 V^*}{\partial {C_W}^2} = \frac{2\,r\,X\,(R_V+v)}{M_U\,R_v\,v\,(X-C_W)^3}$$

which is positive under participation constraint $X - C_W > 0$. The violations of user privacy are increasingly higher as website operational costs rise. In terms of government policy, regulatory interventions to decrease or monitor privacy violations that increase costs to the publisher websites will have an offsetting negative effect on privacy violation. Any positive impacts of intervention must also consider these negative consequences.

For part (v), the first derivative is:

$$\frac{\partial P_W^*}{\partial v} = \frac{M_D\,r^2\,X\,(R_V^2+v^2)}{2\,F\,M_U\,R_V\,v^2\,(X-C_W)}$$

which is positive when $X - C_W > 0$. From this, we observe that the publisher website increases its the price it charges to users (P^W) as user perceptions of privacy violations increase (v). The reason behind this is that as we see in part (i) of Proposition 1, the amount of actual violation decreases with increased privacy concerns which results in lower website profit from third-parties, because less number of third-parties are inclined to participate (equation 3). On the other hand, because of lower number of third-parties in the website, more users are willing to use the website, and the website is better off by increasing the price. The second derivative provides that:

$$\frac{\partial^2 P_W^*}{\partial v^2} = -\frac{M_D\,r^2\,X\,R_V}{F\,M_U\,v^3\,(X-C_W)}$$

which is negative under participation constraint $X - C_W > 0$. This means that the increase in price as publisher website's answer to increased user privacy is greater when privacy concerns are relatively low compared to when privacy concerns are generally heightened. Interpreting this, we can expect to see greater price shocks for users during the initial phases of privacy awareness in the population, as users are jarred from being totally unconcerned about privacy. Subsequent incremental increases in perceived privacy will still increase user prices, but at progressively smaller increments.

Part (vi) is shown by taking the first derivative:

$$\frac{\partial P_W^*}{\partial r} = -\frac{M_D\, r\, X\,(R_V^2 - v^2)}{2\, F\, M_U\, R_V\, v\, (X - C_W)}$$

which is negative under user participation constraint $X - C_W > 0$ and third-party participation constraint $R_V - v > 0$. The first derivative indicates that as the publisher website royalties paid by third-party (r) increases, the prices charged to users will decrease. This is intuitive, as it shows that the website tries to increase the number of third-parties who now pay more royalties, by increasing number of users through reduced prices. Taking the second derivative we have:

$$\frac{\partial^2 P_W^*}{\partial r^2} = -\frac{M_D\, X\,(R_V^2 - v^2)}{F\, M_U\, R_V\, v\, (X - C_W)}$$

which is negative under participation constraint $X - C_W > 0$ and when $R_V - v > 0$. So the decrease in user prices in response to a unit decrease in royalties is greater when royalties are relatively low.

For part (vii), the first derivative is:

$$\frac{\partial P_W^*}{\partial R_V} = -\frac{M_D\, r^2\, X\,(R_V^2 + v^2)}{2\, F\, M_U\, R_V{}^3\, (X - C_W)}$$

which is negative under participation constraint $X - C_W > 0$, noting that the website publisher's equilibrium price to users will decrease if the third-party's revenue from each unit of violation from each user (R_V) increases. With higher profits to third parties for each violation, the website is incentivized to increase the number of users through reduced user pricing. We see this with examples such as Apple offering it's travel mapping and geo-location tracking systems free to users in exchange for extensive personal information on travel behaviour. Taking the second derivative we have:

$$\frac{\partial^2 P_W^*}{\partial R_V{}^2} = \frac{M_D\, r^2\, X\, v}{F\, M_U\, R_V{}^3\, (X - C_W)}$$

which is positive under participation constraint $X - C_W > 0$. This means that the decrease in user price as a result of higher information violation revenue of third parties (R_V) is increasingly larger when violation revenue is higher. For policy makers, this implies that if violation revenues to third parties are currently high, then the initial impact of regulatory imposition of staged reduction over time in third-party violation revenues will result in more significant user price increases early in the regulatory intervention. It also means that if violation revenues to third-parties will increase in the future due to technological innovations that allow third-parties to re-identify website visitors with greater efficiency and effectiveness, then we can only expect the user price shocks that would result from curtailing third-party behaviour only to increase over time.

For Part (viii) we have:

$$\frac{\partial P_W^*}{\partial C_W} = \frac{F\, M_U\, R_V\, v\, (X - C_W)^2 - X\, M_D\, r^2\,(R_V^2 - v^2)}{2\, F\, M_U\, R_V\, v\, (X - C_W)^2}$$

And so the sign of the derivative depends on the sign of numerator. So when we have $F\, M_U\, R_V\, v\, (X - C_W)^2 - X\, M_D\, r^2\, (R_V^2 - v^2) > 0$, the optimal price increases with per user operating costs, and when $F\, M_U\, R_V\, v\, (X - C_W)^2 - X\, M_D\, r^2\, (R_V^2 - v^2) < 0$, it decreases with per user operating costs. This is counterintuitive, as one might think that costs always increase price. However, it is possible for the website to still decrease the prices with increase in costs, and put more emphasis on profit from third-parties. Taking the second derivative we have:

$$\frac{\partial^2 P_W^*}{\partial C_W{}^2} = -\frac{M_D\, X\, r^2\, (R_V^2 - v^2)}{F\, M_U\, R_V\, v\, (X - C_W)^3}$$

which is always negative. This means that the change in optimal price due to increased per user operating costs occurs at decreasing rate.

3.3 Effect of Model Parameters on Number of Users and Third-Parties

By substituting the optimal formula for V^* and P_W^* into equations (5) and (6) optimal number of users and third-parties on website can be obtained. Proposition 2 below provides the optimal number of users and its properties.

Proposition 2.

The optimal number of users on the website N_U^* is calculated as follows:

$$N_U^* = \frac{M_U (X - C_W)}{2\, X} \geq 0$$

It can be seen that for the users to use the website, the participation constraint $X - C_W \geq 0$ must be met. The number of users on the website satisfies the following:

(i) N_U^* is independent of users' perceived privacy violation from each third-party (v).

(ii) N_U^* is independent of publisher website royalties paid by third-party (r), and of third-party's revenue from each unit of violation from each user (R_v).

(iii) N_U^* decreases with website operating costs per user (C_W) at a constant rate.

(iv) N_U^* increases with users' maximum intrinsic value for the website (X) at a decreasing rate.

◊

For part (i) there is no v in the equation for number of users, so the optimal number of users on the website (N_U^*) is independent of the perceived privacy violation (v). While the first part of the proposition seems counter-intuitive at first, it becomes clear when it is looked at from publisher website's perspective, and considering part (i) of Proposition 1. The publisher website changes its privacy violation behavior in a way that it does not sacrifice its users, and thus retains a steady user base. This is an interesting finding which implies that when users are more concerned about their privacy, publisher website reduces the amount of violation that third-parties inflict upon users.

For part (ii), we note again that r and R_V are not present in the formula for optimal number of users, so number of users is independent of these parameters. The number of users is not affected by the third-party royalties and revenue from information violation. This result is not obvious, *a priori*.

For part (iii) we have:

$$\frac{\partial N_U^*}{\partial C_W} = -\frac{M_U}{2X}$$

This shows that there will be fewer users of publisher websites when the per user operational costs of the website increase.

For part (iv) we have:

$$\frac{\partial N_U^*}{\partial X} = \frac{M_U C_W}{2X^2}$$

The number of users increases with their maximum intrinsic value for the website. This is an expected result, because if users generally have a higher potential value for the website, they are more willing to use it. By taking the second derivative we have:

$$\frac{\partial^2 N_U^*}{\partial X^2} = -\frac{M_U C_W}{X^3}$$

This means that the increase in optimal number of users on the website occurring as a result of increased valuation occurs at a decreasing rate as the valuation increases.

Next, Proposition 3 provides the optimal number of third-parties and its properties.

Proposition 3.

The optimal number of third-parties in the website (N_D^*) *is calculated as follows:*

$$N_D^* = \frac{M_D r (R_V - v)}{2Fv} \geq 0$$

It can be seen that for the third-parties to join the website, the participation constraint $R_V - v \geq 0$ *must be met. The number of third-parties in the website satisfies the following:*

(i) N_D^* *decreases with users' perceived privacy violation from each third-party* (v) *at an increasing rate.*

(ii) N_D^* *increases with publisher website royalties paid by third-party* (r).

(iii) N_D^* *increases with third-party's revenue from each unit of violation from each user* (R_V).

(iv) N_D^* *decreases with third-parties maximum fixed cost* (F) *at an increasing rate.*

◊

For part (i) the first derivative is as follows:

$$\frac{\partial N_D^*}{\partial v} = -\frac{M_D r R_V}{2Fv^2}$$

The first derivative is always negative, meaning that number of third-parties is decreasing in v. This is a result of lower violation factors set by the publisher website for higher user privacy concerns, as shown in part (i) of Proposition 1. Additional third-parties pose additional risks to violation and misuse of user and website information. Thus, websites dealing with sensitive user information would tend to do most of their tasks themselves, rather than engaging third-parties to perform needed tasks. Taking the second derivative we have:

$$\frac{\partial^2 N_D^*}{\partial v^2} = \frac{M_D\, r\, R_V}{F\, v^3}$$

The second derivative is always positive, meaning that the rate at which number of third-parties decrease with increased users' perception of privacy violations is increasing when user concerns are higher. Thus, publisher websites are especially sensitive to high values of users' perceptions of privacy violation, and are willing to drastically reduce amount of violation when already heightened perceptions of privacy violation increase still further.

This is a particularly interesting result, partly because it is observable. Given this result, we would expect websites dealing with sensitive content and data to deal with substantially fewer third parties, compared to sites where no sensitivity exists for the user. If we can observe this effect in practice, then this would lend considerable validation to the model.

For part (ii) we have

$$\frac{\partial N_D^*}{\partial r} = \frac{M_D\,(R_V - v)}{2\,F\,v}$$

which is positive when $R_V - v > 0$. This is expected, as an increase in royalties will reduce third-party profit and reduces their participation in publisher website.

For part (iii) we have

$$\frac{\partial N_D^*}{\partial R_V} = \frac{M_D\, r}{2\,F\,v}$$

which is always positive. This is also an expected result that states that there are more third-parties willing to join the website when the third-parties' per user revenue per unit of violation is higher. This is because more third-parties are willing to join the publisher website due to higher profits.

For part (iv) we have:

$$\frac{\partial N_D^*}{\partial F} = -\frac{M_D\, r\,(R_V - v)}{2\,F^2 v}$$

which is negative as long as the participation constraint $R_V - v > 0$ is met. Intuitively, the number of third-parties on the website decreases as their fixed cost increases. Taking the second derivative we have:

$$\frac{\partial^2 N_D^*}{\partial F^2} = \frac{M_D\, r\,(R_V - v)}{F^3 v}$$

which is always positive, meaning that for high maximum fixed cost values, the decrease in number of third-parties as a result of increase in fixed cost is exacerbated when fixed cost are high compared to low.

4. Results and Analysis

In this section, we discuss the results and implications of the model. We first show how the model predicts the effects of privacy considerations on industry concentration in the third-party market. Next, the social effects of privacy concerns are analyzed, and then the implications for policy-making are discussed.

4.1 Effect of Privacy Concerns on Market Concentration

Here we show how the market concentration is affected by the privacy concerns of the users. We consider two cases: 1. when all of the third-parties have homogenous shares of the market, and 2. when third-parties can have non-homogenous shares of the market. The first case is used to infer the second. We use the Herfindahl-Hirschman Index (HHI) as a recognized measure for market concentration. The HHI is calculated as follows:

$$HHI = \sum_{i=1}^{N_D} s_i^2 \tag{9}$$

where s_i is the market share of ith firm, or in this case, third-party.

4.1.1 Third-Parties with Homogenous Market Shares

When all the third parties have equal share of the market, the market share of each third-party is simply calculated as $s_i = 1/N_D$ and the HHI is calculated as:

$$HHI = \sum_{i=1}^{N_D} (1/N_D)^2 = N_D(1/N_D)^2 = 1/N_D \tag{10}$$

By inserting the optimal number of third-parties from Proposition 3 , we have:

(Md r (Rv (1 + tPw) - (1 + tr) v))/(2 F v)

$$HHI = 1/N_D == 1/[M_D \frac{r\,(R_V - v)}{2\,F\,v}] = \frac{2\,F\,v}{M_D\,r\,(R_V - v)} \tag{11}$$

Taking the first derivative of HHI with respect to v we have:

$$\frac{\partial\,HHI}{\partial\,v} = \frac{2\,F\,R_V}{M_D\,r\,(R_V - v)^2}$$

which is always positive, so HHI is increasing in v, or the market concentration is increasing in the user privacy concern.

Other than the factors affecting the number of third-parties as provided in formulae for N_D above, barriers to entry also affect the number of third-parties. Higher barriers result in less third-parties (by definition) and less concentration (Mueller and Hamm 1974). To include the effect of barriers to entry, we rewrite the total number of potential third-parties, M_D to be as M_D/B, where B is the level of barrier. This means that higher level barriers will reduce the number of potential third-parties. Now we can rewrite the HHI formula as:

$$HHI = 1/N_D == 1/[(M_D/B)\,\frac{r\,(R_V - v)}{2\,F\,v}] = \frac{2\,FB\,v}{M_D\,r\,(R_V - v)} \tag{12}$$

It can be seen that HHI is increasing in the entry barrier level, so the market concentration is increasing in the level of barrier to entry. The level of barrier to entry is higher for the third-parties that operate in areas with high privacy concerns and high information sensitivity. In practice, a reason for this is that these third-parties need to invest more in their information technology security, which is associated with higher sunk entry cost, and entry cost is a major barrier to entry.

Next we will generalize the results to the case where third-parties have non-homogenous market shares.

4.1.2 Third-Parties with Non-Homogeneous Market Shares

While previously we assumed the market shares to be homogenous for all third-parties, this is not realistic in most cases. In the markets with higher information sensitivity and higher risks for privacy violations, it is logical for publisher websites to do more research in selecting the third-parties, and it is possible that they tend to incorporate prominent third-parties rather than smaller ones. This will result in a market with non-homogenous market shares. Here we include this non-homogeneity in our calculations.

Going back to the HHI formula in (9), it can be seen that the lowest value for HHI is achieved for when the market shares for third-parties are all equal. This is because the sum of squares of variables, when the sum of variables is fixed, is minimized when the variables are all equal. So we can make an argument that the market concentration is lowest for the case of homogenous market shares, and having different market shares for third-parties will increase the HHI. Also, knowing that higher information sensitivity will increase the barrier to entry level, this will result in even more diverse market shares for third-parties. So we can state that market concentration is even higher in case of non-homogenous third-party market shares compared to the homogeneous case.

4.2 Effect of Users' Perception of Privacy Violation on Stakeholders

In this section we discuss how the model parameters affect the sharing behavior, and how market interventions by government and policy-makers can affect the surpluses of users and third-parties, website profit, and eventually social welfare.

Here we consider how the users' perception of privacy violation, v, can affect the stakeholders. The perception of privacy can be increased by education through media, or by increased website transparency. An important public policy question is what happens when the users' perception of violation is imperfect, that is when privacy concerns are too high or too low. Here we study the effect of these manipulations in v. Proposition 4 provides the effect of model parameters on stakeholders and society.

Proposition 4.

When the publisher website chooses V^* and P_W^* so as to maximize profit, website profit Π_W^*, user surplus CS^*, third-party surplus DS^*, and social welfare SW^* satisfy the following:

(i) Π_W^* *decreases with users' perceived privacy violation from each third-party* (v) *at an increasing rate.*

(ii) CS^* *is independent of users' perceived privacy violation from each third-party* (v)*.*

(iii) DS^* *decreases with users' perceived privacy violation from each third-party* (v)*, at an increasing rate.*

(iv) SW^* *decreases with users' perceived privacy violation from each third-party* (v)*, and it decreases at an increasing rate when* $3\,F\,R_V + 2\,M_D v - 2\,F\,v > 0$.

◊

For part (i), note that the optimal profit of the publisher website can be calculated using equation (7) and substituting the optimal values of N_U, N_D, and P_W, which yields:

$$\Pi_W^* = \frac{F\,M_U\,R_V\,v\,(C_W^2 + X^2) + (\,M_D\;r^2\,(\,R_V - v)^2 - 2\,F\,(2\,F_W + C_W\,M_U)\,R_V v\,)\,X}{4\,F\,R_V v\,X}$$

Taking the first derivative of Π_W^* with respect to v we have:

$$\frac{\partial \Pi_W^*}{\partial v} = -\frac{M_D\,r^2\,(R_V^2 - v^2)}{4\,F\,R_V\,v^2}$$

which is negative when $R_V - v \geq 0$. This is intuitive, because when users become concerned about their privacy, website cannot monetize user information through third-parties, and its profit decreases. Taking the second derivative we have:

$$\frac{\partial^2 \Pi_W^*}{\partial v^2} = \frac{M_D\,r^2\,R_V}{2\,F\,v^3}$$

which is positive, meaning that the reduction in profit as a result of higher user perception of privacy occurs at an increasing rate as user privacy concerns rise.

For part (ii) we first need to calculate user surplus, defined as the sum of the utility of all users with positive utility:

$$CS = \int_{\hat{x} = P^W + N_D v\,V}^{X} (x - N_D v\,V - P^W)\,dx = \tfrac{1}{2}(X - N_D v\,V - P^W)^2$$

By substituting the optimal values for N_D, P^W, and V in the user surplus formula above, we can calculate the optimal user surplus as:

$$CS^* = \tfrac{1}{8}(X - C_W)^2$$

It can be seen that the user surplus does not depend on users' perception of privacy concern. The reason for this is that as the users' perception of privacy concern (v) changes, the publisher website adjusts the actual violation factor (V) so that the level of perception combined with actual violation ($v\,V$) remains constant. This implies that there is no incentive for the users to improve their perception of privacy.

For part (iii) we need to calculate the third-party surplus, defined as the sum of profits of all third-parties available on the publisher website:

$$DS = \int_0^{\hat{f}=N_U V R_V - r} (N_U V R_V - r - f)\, df = \tfrac{1}{2}(N_U V R_V - r)^2$$

By substituting the optimal values for N_U and V in the third-party surplus formula above, we can calculate the optimal third-party surplus as:

$$DS^* = \frac{r^2 (R_V - v)^2}{8 v^2}$$

Taking the first derivative of third-party surplus with respect to v we have:

$$\frac{\partial DS^*}{\partial v} = -\frac{r^2 R_V (R_V - v)}{4 v^3}$$

which is negative when $R_V - v > 0$, meaning that the third-party surplus decreases with users' perception of violation. This is intuitive, because when the users' perception of violation from third-parties increases, the actual violation factor (V) is decreased by the publisher website, and this decreases the third-party profits. Taking the second derivative we have:

$$\frac{\partial^2 DS^*}{\partial v^2} = \frac{r^2 R_V (3R_V - 2v)}{4 v^4}$$

which is positive when $3R_V - 2v > 0$, and is true when $R_V - v > 0$. This means that the rate at which the third-party surplus decreases with users' perception of privacy is increasing. In other words, the drop in third-party surplus is more significant for larger values of v.

Part (iv) is straightforward, as the social welfare is sum of the publisher website profit, user surplus, and third-party surplus:

$$SW = CS + DS + \Pi_W$$

Taking the derivative with respect to v:

$$\frac{\partial SW^*}{\partial v} = -\frac{r^2 (R_V - v)(F R_V^2 + M_D v (R_V + v))}{4 F R_V v^3}$$

which is negative when $R_V - v > 0$, and so the social surplus is decreasing with v. For the second derivative we have:

$$\frac{\partial^2 SW^*}{\partial v^2} = \frac{r^2 R_V (3 F R_V - 2 F v + 2 M_D v)}{4 F v^4}$$

which is positive when $3 F R_V + 2 M_D v - 2 F v > 0$, meaning that the rate at which social welfare decreases with v is increasing in this case.

In summary, users may have irrationally high privacy concerns due to media hype, or if there seems to be no user protection against future use of data such as data re-identification, abuse, law enforcement use of data, data leakage (data going into undesirable or unauthorized entities). As seen in Proposition 4, this is harmful to all of the stakeholders, except for users who are indifferent. In case of low perceived privacy concerns, the user data is being used without user's awareness. While the effect of user perception is negative for third-party and website and has no effect on user surplus in our model, it should be noted that the violations can result in catastrophic events of information being leaked to inappropriate entities, which could hugely impact the publisher website.

4.3 Implications of the Model For Policy-Makers

Here we consider the effect of regulatory organizations that can force certain taxes. We model this by performing the following transformations on the original model. The changes are made to the user utility equation (1), and third-party profit equation (3).

$$P_W \to P_W(1+T_{P_W}), \qquad -1 < T_{P_W} < 1 \tag{13}$$

$$r \to r(1+T_r), \qquad -1 < T_r < 1 \tag{14}$$

where T_{P_W} and T_r are the taxations set by an outside organization or government on website price and third-party royalties paid to the website, respectively. Note that the taxation can represent risk reserves that might be set by policy-makers to be used in case that an adverse incident takes place, or can be seen as setting standards on information handling that might make the transactions harder and more costly. The taxations can also take negative values, meaning that the policy-maker takes action to make the processes easier or less costly. Lemma 2 provides the optimal information violation factor and publisher website price when these taxations are included in the base model.

Lemma 2.

The platform monopolist chooses

$$V^* = \frac{r\,X\,(R_V(1+T_{P_W})+v(1+T_r))}{M_U\,R_v(X-C_W(1+T_{P_W}))}$$

$$P_W^* = \frac{F\,M_U\,R_v\,v\,(X^2-C_W^2\,(1+T_{P_W})^2)-X\,M_D\,r^2\,(R_V^2(1+T_{P_W})^2-v^2(1+T_r)^2)}{2\,F\,M_U\,R_v\,v\,(1+T_{P_W})\,(X-C_W(1+T_{P_W}))} \geq 0$$

◊

The proof for this Lemma follows from the proof for Lemma 1 and by making the transformations (13) and (14) in user utility equation (1), and third-party profit equation (3).

Using the Lemma 2, Propositions 5 below provides the effect of different parameters on the publisher website decision variables of optimal violation factor V and website price P_W. Note that in this model, the user participation constraint changes to $X - C_W(1+T_{P_W}) > 0$ and the third-party participation constraint remains the same ($R_V - v \geq 0$).

Proposition 5.

The optimal violation factor V^* *and the optimal website price* P_W^* *for the case with taxations satisfy the following:*

(i) V^* *increases with taxation on publisher website price* (T_{P_W}) *at an increasing rate.*

(ii) V^* increases with taxation on third-party royalties (T_r).

(iii) P_W^* can increase or decreases with taxation on publisher website price (T_{P_W}) at an increasing rate.

(iv) P_W^* increases with taxation on third-party royalties (T_r) at a constant rate.

◊

For part (i) we have:

$$\frac{\partial V^*}{\partial T_{P_W}} = \frac{r\,X\,(R_V\,X + C_W\,v\,(1+T_r))}{M_U\,R_V\,v\,(X - C_W(1+T_{P_W}))^2}$$

which is positive when $X - C_W(1+T_{P_W}) > 0$, meaning that the website will increase the amount of violation as a response to taxation on prices. This is intuitive, as the taxation will reduce the users utility for the website and reduce website profit from third-parties. The publisher website will try to increase its profit by increasing user information monetization. The second derivative is calculated as follows:

$$\frac{\partial^2 V^*}{\partial {T_{P_W}}^2} = \frac{2\,C_W\,r\,X\,(R_V\,X + C_W\,v\,(1+T_r))}{M_U\,R_V\,v\,(X - C_W(1+T_{P_W}))^3}$$

which is positive when $X - C_W(1+T_{P_W}) > 0$, meaning that the increase in V^* as a result of taxation on website price is higher for more extreme taxation rates.

As for part (ii) the first derivative is calculated as:

$$\frac{\partial V^*}{\partial T_r} = \frac{r\,X}{M_U\,R_V\,(X - C_W(1+T_{P_W}))}$$

which is positive when $X - C_W(1+T_{P_W}) > 0$, meaning that the publisher website will increase the amount of violation (V^*) as a result of taxation on third-party. Similarly to part (i) of this proposition, as a result of reduced third-party interest because of taxation, publisher website increases third-party motivation by increasing the user privacy violation.

For part (iii) the first derivative is calculated as:

$$\frac{\partial P_W^*}{\partial T_{P_W}} = -\frac{F\,X\,M_U\,R_V\,v\,(X - C_W(1+T_{P_W}))^2 + X^2\,M_D\,r^2\,(R_V^2(1+T_{P_W})^2 - v^2(1+T_r)^2) - 2\,M_D\,X\,C_W\,(1+T_{P_W})\,v^2\,r^2\,(1+T_r)^2}{2\,F\,M_U\,R_V\,v\,(1+T_{P_W})^2(X - C_W(1+T_{P_W}))^2}$$

So the sign of the derivative depends on the sign of numerator, and can be increasing or decreasing. For the second derivative we have:

$$\frac{\partial^2 P_W^*}{\partial {T_{P_W}}^2} = \frac{M_D\,r^2\,v\,(1+T_r)^2 + F\,M_U\,R_V X}{2\,F\,R_V\,(1+T_{P_W})^3}$$

Which is always positive, noting that the change in optimal website price occurs at increasing rate.

For part (iv) the first derivative is calculated as:

$$\frac{\partial P_W^*}{\partial T_r} = \frac{M_D r^2 v X (1+T_r)}{F M_U R_V (1+T_{P_W}) (X - C_W (1+T_{P_W}))}$$

which is positive when $X - C_W(1+T_{P_W}) > 0$, meaning that the website will increase its website price as taxation on third-parties increases. The second derivative is calculated as

$$\frac{\partial^2 P_W^*}{d T_r^2} = \frac{M_D r^2 v X}{F M_U R_V (1+T_{P_W}) (X - C_W (1+T_{P_W}))}$$

which is positive when $X - C_W(1+T_{P_W}) > 0$, meaning that the increase in website price as a result of taxation on third-parties is more significant when taxation is higher.

Next, we consider the effect of taxation on stakeholders. First, we look at the taxation on publisher website price.

Proposition 6.

When the publisher website chooses V^* and P_W^* so as to maximize profit, when taxation on users is possible, website profit Π_W^*, user surplus CS^*, third-party surplus DS^*, and social welfare SW^* satisfy the following:

(i) Π_W^* decreases with taxation on publisher website price (T_{P_W}) at an increasing rate.

(ii) CS^* decreases with taxation on publisher website price (T_{P_W}) at an increasing rate.

(iii) DS^* increases with taxation on publisher website price (T_{P_W}) when $R_V(1+T_{P_W}) - v(1+T_r) > 0$ at an increasing rate.

(iv) SW^* can increase or decrease in taxation on publisher website price (T_{P_W}) at an increasing rate.

◊

For part (i) , note that the optimal profit of the publisher website can be calculated using profit equation and substituting the optimal values of N_U, N_D, and P_W, which yields:

$$\Pi_W^* = \frac{F M_U R_V v (C_W^2 (1+T_{P_W})^2 + X^2) + (M_D \; r^2 (R_V (1+T_{P_W}) - v(1+T_r))^2 - 2 F (2 F_W + C_W M_U) R_V (1+T_{P_W}) v) X}{4 F R_V v X (1+T_{P_W})^2}$$

the first derivative is calculated as:

$$\frac{\partial \Pi_W^*}{\partial T_{P_W}} = \frac{X M_D r^2 (R_V{}^2 (1+T_{P_W})^2 - v^2 (1+T_r)^2) - F M_U R_V v (X^2 - C_W^2 (1+T_{P_W})^2)}{4 F R_V v X (1+T_{P_W})^2}$$

The sign of the derivative is negative, because of the assumption of $P_W{}^* \geq 0$ in Lemma 2 which states:

$$\frac{F M_U R_v v (X^2 - C_W{}^2 (1+T_{P_W})^2) - X M_D r^2 (R_V{}^2 (1+T_{P_W})^2 - v^2 (1+T_r)^2))}{2 F M_U R_v v (1+T_{P_W}) (X - C_W (1+T_{P_W}))} \geq 0$$

$$\Rightarrow F\,M_U\,R_v\,v\,(X^2 - C_W^{\,2}\,(1+T_{P_W})^2) - X\,M_D\,r^2\,(R_V^{\,2}(1+T_{P_W})^2 - v^2(1+T_r)^2\,)\,) \ge 0$$

$$\Rightarrow X\,M_D\,r^2\,(R_V^{\,2}(1+T_{P_W})^2 - v^2(1+T_r)^2\,)\,) - F\,M_U\,R_v\,v\,(X^2 - C_W^{\,2}\,(1+T_{P_W})^2) \le 0$$

$$\Rightarrow \frac{X\,M_D\,r^2\,(R_V^{\,2}(1+T_{P_W})^2 - v^2(1+T_r)^2\,) - F\,M_U\,R_V\,v\,(X^2 - C_W^2(1+T_{P_W})^2)}{4\,F\,R_V v\,X\,(1+T_{P_W})^2} = \frac{\partial \Pi_W^*}{\partial T_{P_W}} \le 0$$

So the profit of the publisher website is decreasing in taxation on website price. For the second derivative we have:

$$\frac{\partial^2 \Pi_W^*}{\partial T_{P_W}^{\;2}} = \frac{M_D\,r^2\,v\,(1+T_r)^2 + F\,M_U\,R_V X}{2\,F\,R_V(1+T_{P_W})^3}$$

Which is always positive, meaning that the decrease in profit due to taxation on website price, is higher for larger taxation values.

For part (ii) the user surplus is calculated as follows:

$$CS^* = \tfrac{1}{8}(X - C_W(1+T_{P_W}))^2$$

Taking the first derivative with respect to T_{P_W} we have:

$$\frac{\partial CS^*}{\partial T_{P_W}} = -\frac{C_W(X - C_W(1+T_{P_W}))}{4}$$

which is negative when $X - C_W(1+T_{P_W}) > 0$, meaning that user surplus also decreases with taxation on website price. This is intuitive, as we expect that the taxation to reduce the user utility, and thus reduce the total user surplus. Taking the second derivative we have:

$$\frac{\partial^2 CS^*}{\partial T_{P_W}^{\;2}} = \frac{C_W^{\;2}}{4}$$

which is always positive, noting that the decrease in user surplus as a result of taxation on third-party price is higher for larger taxation values.

For part (iii), the third-party surplus is calculated as:

$$DS^* = \frac{r^2\,(R_V(1+T_{P_W}) - v(1+T_r))^2}{8\,v^2}$$

Taking the first derivative with respect to T_{P_W} we have:

$$\frac{\partial DS^*}{\partial T_{P_W}} = \frac{r^2\;R_V\,(R_V(1+T_{P_W}) - v(1+T_r))}{4\,v^2}$$

which is positive when $R_V(1+T_{P_W}) - v(1+T_r) > 0$, meaning that the third-party surplus is increasing in taxation on publisher website price. The reason for this is that when there is taxation on publisher website price, the website will increase the violation factor (Proposition 5 part (i)), which will benefit the third party, and this benefit is higher than the loss due to lower number of users in the website. Taking the second derivative we have:

$$\frac{\partial^2 DS^*}{\partial T_{P_W}{}^2} = \frac{r^2\ R_V{}^2}{4\,v^2}$$

which is always positive, noting that the increase in third-party surplus due to taxation on third-party is even higher for higher taxation values.

For part (iv) the social welfare is calculated as the sum of the publisher website profit, user surplus, and third-party surplus:

$$SW = CS + DS + \Pi_W$$

The first derivative is calculated as

$$\frac{\partial SW^*}{\partial T_{P_W}} = -\frac{1}{4FR_V v^2 X(1+T_{P_W})^2}$$

$$(C_W F R_V v^2 X^2 - C_W^2 F R_V v^2\,(1+T_{P_W})^3\,X - XM_D\,r^2\,v\,(R_V{}^2(1+T_{P_W})^2 - v^2(1+T_r)^2\,) - FR_V{}^2 r^2\,(1+T_{P_W})^2(R_V\,(1+T_{P_W}) - v\,(1+T_r)\,)$$

$$FR_V M_U v^2(C_W^2(1+T_{P_W})^2 + X^2))$$

The sign of the derivative depends on the sign of the numerator, which can be positive or negative. Taking the second derivative we have:

$$\frac{\partial^2\,SW^*}{\partial T_{P_W}{}^2} = \frac{2\,M_D\,r^2(1+T_r)^2 + F\,R_V(r^2\,R_V{}^2(1+T_{P_W})^3 + v^2(C_W^2(1+T_{P_W})^3 + 2M_U X))}{4FR_V v^2(1+T_{P_W})^3}$$

which is always positive, meaning that the change in social welfare due to taxation on website price occurs at increasing rate.

Next, we study the effect of third-parties taxations on stakeholders.

Proposition 7.

When the publisher website chooses V^* and P_W^* so as to maximize profit, when taxation on third-party is possible, when having $R_V(1+T_{P_W}) - v(1+T_r) > 0$, website profit Π_W^*, user surplus CS^*, third-party surplus DS^*, and social welfare SW^* satisfy the following:

(i) Π_W^* decreases with taxation on third-party (T_r) at an increasing rate.

(ii) CS^* is independent of taxation on third-party (T_r).

(iii) DS^* decreases with taxation on third-party (T_r) at an increasing rate.

(iv) SW^* decreases with taxation on third-party (T_r) at an increasing rate..

◊

Noting the formula for Π_W^*, CS^*, DS^*, and SW^* from Proposition 6, the proofs are as follows. For part (i) the first derivative is calculated as:

$$\frac{\partial \Pi_W^*}{\partial T_r} = -\frac{M_D\,r^2\,(R_V\,(1+T_{P_W}) - v\,(1+T_r)\,)}{2\,F\,R_V\,(1+T_{P_W})}$$

which is negative when $R_V(1+T_{P_W})-v(1+T_r)>0$, meaning that the website profit is decreasing in taxation on third-party. This is intuitive, because the taxation on third-party reduce the third-party profit, and less third-parties are willing to join the website and this decreases website profit. Taking the second derivative, we have:

$$\frac{\partial^2 \Pi_W^*}{\partial {T_r}^2} = \frac{M_D\, r^2\, v}{2\, F\, R_V\, (1+T_{P_W})}$$

Which is always positive, meaning that the decrease in profit as a result of taxation on third-parties is higher for higher taxation values.

For part (ii) we can see that the formula for CS^* does not entail T_r, and is thus independent of it. This is because while the taxation affects number of third-parties, the publisher website increases violation factor, so that the total violation done to the users remains the same.

For part (iii) we have:

$$\frac{\partial DS^*}{\partial T_r} = -\frac{r^2\, (R_V\, (1+T_{P_W})\, - v\, (1+T_r)\,)}{4\, v}$$

Which is negative under the participation constraint, meaning that the third-party surplus is also decreasing in third-party taxation. This is intuitive, as the taxation focuses the third-party, reducing their profit and thus surplus. Taking the second derivative we have:

$$\frac{\partial^2 DS^*}{\partial {T_r}^2} = \frac{r^2}{4}$$

which is always positive, meaning that the decrease in third-party surplus due to third-party taxation is even higher for higher taxation values.

For part (iv) we have:

$$\frac{\partial SW^*}{\partial T_r} = -\frac{r^2\, (F\, R_V\, (1+T_{P_W}) + 2\, M_D\, v\,)\, (R_V\, (1+T_{P_W}) - v\, (1+T_r)\,)}{4\, F\, R_V v\, (1+T_{P_W})}$$

which is again, always negative, so the total social welfare decreases with taxation on third-parties. The indication for this is that the taxation on third-parties is actually not good for the the stakeholders. The second derivative is calculated as:

$$\frac{\partial^2 SW^*}{\partial {T_r}^2} = \frac{r^2}{4}\left(1 + \frac{2\, M_D\, v}{F\, R_V (1+T_{P_W})}\right)$$

which is always positive, noting that the decrease in social welfare due to third-party taxation is higher for higher taxation values.

5. Model Validation

In this section we detail the exploratory analysis performed to analyze the network between publisher websites and third-parties. Alexa Internet provides rankings for websites within 17 different categories, which we call subject categories. We carry out exploratory validation study on the 100 most-visited publisher websites from seven of these 17 subject categories (news, arts, shopping, kids and teens, health, business, and adult) provided and ranked by Alexa website rankings[1]. These seven categories were selected with the intention of finding subject categories for which users might reasonably be expected to have different intentions to disclose personal information and browsing behaviour due to the nature of the subject content. The study was conducted as follows. An automated browser accesses a publisher website's home page, and the connections that are made from publisher website to third-parties are observed. We use page loading time plus a 3-second window to collect data gathered using a residential internet plan and using Lightbeam for Firefox (Windows) to record these connections.

To better capture the structure of the industry, we profile the third-parties and separate them based on the industry sectors as classified by Cookiepedia.co.uk. The three industry sectors are targeting/advertising (T/A), functionality (F), and performance (P). For those third-parties that are not profiled in Cookiepedia.co.uk, we make a judgment using available information. 1893 third-party websites are identified in total, with 568 T/A, 487 F, 627 P, and 211 classified as unknown (U). Using different domain finder services[2], multiple third-party websites in each sector owned by the same company are treated as a single third-party for analysis, entailing 1066 unique owner companies comprising 442 T/A, 336 F, and 340 P, with some owner companies providing services in multiple categories. The number of connections made and number of cookies used follow a similar pattern to number of third-parties, and so we provide the analysis based on number of third-parties only.

5.1 Observations

Noting that information sensitivity and user privacy concerns likely varies among different websites, we expect the sharing behavior to differ for websites with different subjects. Figure 2 provides the sharing behavior for the top 100 publisher websites in each subject category and industry sector.

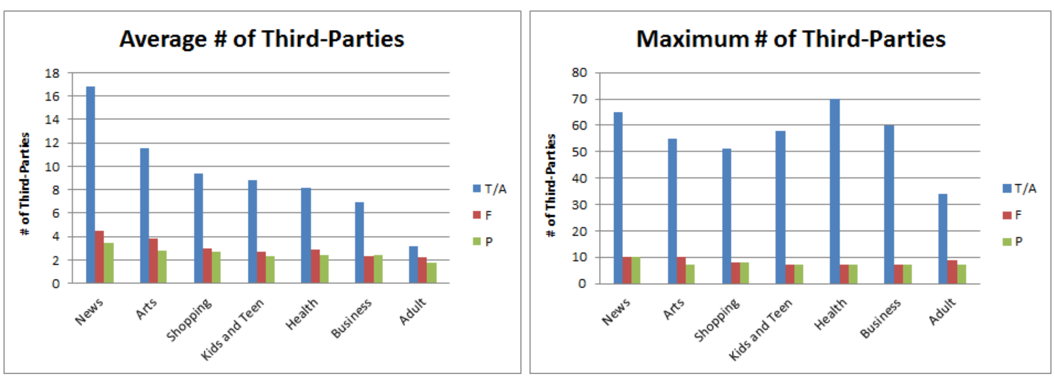

Figure 2. Third-Party Sharing by Subject Categories and Industry Sectors

The number of third-parties used in the T/A industry sector is significantly higher than for both F and P. The F sector has significantly higher sharing than P in 3 of the subject categories and overall. The T/A, F, and P sectors comprise 60, 20, and 15% of all third-party connections, with 5% unknown (U). Thus, sharing across different third-party sectors varies.

[1] Alexa.com/topsites/category. The Alexa list of website categories is consistent with the Open Directory Project categories found at rdf.DMOZ.org/rdf/categories.txt.

[2] This paper uses whois.domaintools.com, whois.net, and who.is.

The results suggest that user information is falling into the hands of many companies who use the information for advertising purposes which provides credence to concerns raised in the media. In terms of publisher website subject categories, news is the "most shared" category. Note that the publisher website's business model is beyond the scope of this study, but this may also influence the use of third parties. In the case of news, the industry has a history of revenues coming from both advertising and subscription fee business models, and in many cases, companies such as the Los Angeles Times and the New York Times employ both business models simultaneously. The adult subject category has the least average sharing, followed by business and health for all three industry sectors (T/A, F, and P). Specifically, health websites entail sensitive information that is a point of concern for users (Alonso-Zaldivar and Gillum 2015). These observations supports the findings of the model that states sharing is lower for contexts where users' privacy concerns are greater.

We also examine the third-party market concentration measure, using Herfindahl-Hirschman index (HHI) based on publisher websites' average monthly unique visitors in the United States for a single year period ending in March 2014 as provided by compete.com. The T/A sector has the lowest HHI concentrations, followed by P, and then by F for all subject categories analyzed separately. The HHI results are provided in Figure 3. Results indicate that T/A has the least industry concentration. There are relatively fewer third-parties in F and P, and the HHI's indicate that these industry sectors have higher number of large dominant players. The market concentration results are also in tandem with findings of the model, where categories with higher privacy concerns are found to have higher HHI.

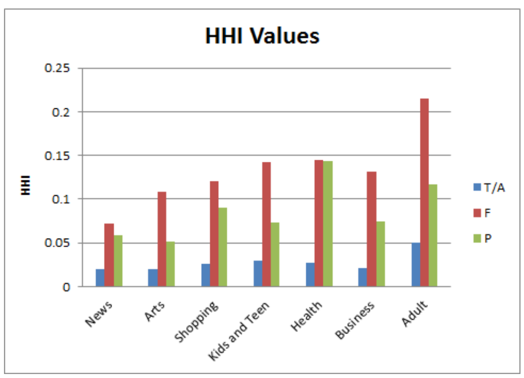

Figure 3. HHI for Third-Party Companies by Subject Categories and Industry Sectors

We do not find evidence for the popularity of publisher websites (as measured by monthly unique visitors) to have any significant effect on third-party sharing.

5.2 Comparing Actual Third-Party Sharing to the Very Low Disclosure Case

We first make the case that active privacy tools such as Adblock Plus enforce the third-party sharing behaviour of the user with very low intention to disclose personal information and browsing behaviour with third parties. Without an active privacy tool, we argue that the website is free to share with the number of third parties that is optimal for their situation, balancing the issues outlined in the model regarding monetization versus maintaining a strong user base. As a control, we test for the impact of the passive privacy "Do Not Track" (DNT) feature, and found that this had no significant effect on average publisher website behaviour with respect to third-party sharing. Thus, we have two conditions of testing for comparison: with and without the use of Adblock Plus. Our analysis is using Adblock plus with the default settings (without "EasyPrivacy" blocking list which blocks user tracking). Adblock Plus represents the very low disclosure (very high privacy) case with maximum benefits for each publisher website.

Using DNT as a control, and Adblock Plus and no privacy tool as our two test conditions, we compare the number of third-parties when browsing with and without these features individually. Figure 4 depicts the total number of third-parties engaged over all 100 publisher websites for each industry sector and then by each subject category. Figure 5 provides the average and maximum number of third-parties used by publisher websites for each industry sector and subject category.

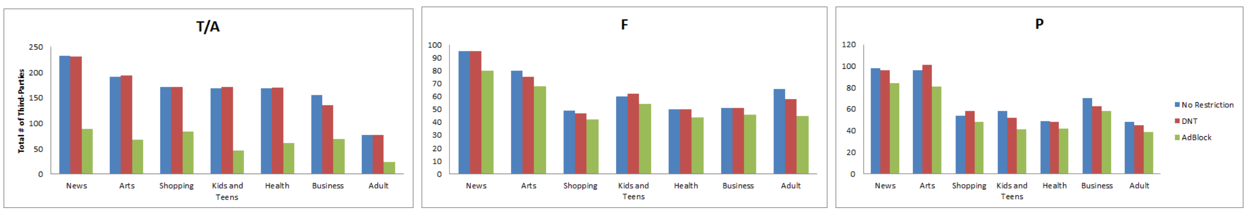

Figure 4. Impact of Privacy Tools on Third-Party Sharing
Combined Over Top 100 Publisher Websites

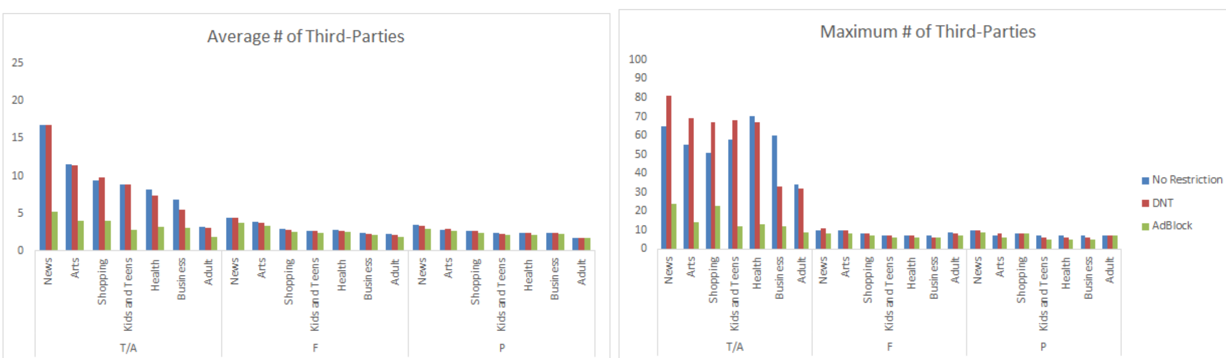

Figure 5. Impact of Privacy Tools on Average and Maximum Third-Party Sharing

Table 3 provides differences in means and P-values for testing whether or not the means are different. A positive difference in means indicates that the average number of third-parties for the column condition (e.g., DNT or AdBlock) is lower than the average for the row condition (e.g., No Restriction or DNT), and vice-versa for negative values. At significance level of 0.01, the average level of third-party sharing does not change when DNT is activated (comparing No Restriction to DNT) for any industry sector/subject category combination. Thus, we find no evidence for passive privacy tools impacting third-party sharing. The opposite is true for Adblock Plus. Comparing No Restriction to Adblock Plus, significant differences in T/A were found in all subject categories except for adult. Significant differences in F and P were found in news, arts, kids and teens, and healthcare, but not for shopping, business, and adult. Comparing DNT to Adblock Plus, results are the same for T/A as when comparing No Restriction to Adblock Plus (significant differences in T/A were found in all subject categories except for adult). Significant differences comparing DNT to Adblock Plus for F and P were found in news, arts, shopping, kids and teens, and healthcare, but not for business, and adult. Adblock Plus, overall (combining T/A, F, P, and U), significantly reduces the number of third-parties in all subject categories compared to both No Restriction and DNT, whereas DNT has no differences in means compared to No Restriction.

Table 3. Impact of Privacy Tools on Information Sharing

		T/A				F				P				All (Including U)			
		Difference in Means		P-Value		Difference in Means		P-Value		Difference in Means		P-Value		Difference in Means		P-Value	
		DNT	AdBlock	DNT	AdBlock	DNT	AdBlock	DNT	AdBlock	DNT	AdBlock	DNT	AdBlock	DNT	AdBlock	DNT	AdBlock
News	NONE	0.3627	11.8922	0.7072	0.000*	0.0882	0.7451	0.3683	0.000*	0.0686	0.5490	0.4088	0.000*	0.3627	12.9510	0.7390	0.000*
	DNT	-	11.8824	-	0.000*	-	0.7451	-	0.000*	-	0.5098	-	0.000*	-	13.1078	-	0.000*
Arts	NONE	0.1635	7.0481	0.7856	0.000*	0.1538	0.5769	0.1483	0.000*	0.0481	0.3173	0.6433	0.001*	0.1538	7.6538	0.8257	0.000*
	DNT	-	7.1058	-	0.000*	-	0.5577	-	0.000*	-	0.3462	-	0.001*	-	7.9038	-	0.000*
Shopping	NONE	-0.2200	5.0500	0.8828	0.000*	0.0900	0.3300	0.7221	0.163	0.0200	0.2500	0.9350	0.287	-0.3600	5.6700	0.8429	0.000*
	DNT	-	5.2700	-	0.000*	-	0.2400	-	0.001*	-	0.2300	-	0.004*	-	6.0300	-	0.000*
Kids and Teens	NONE	-0.1165	5.7670	0.8585	0.000*	0.1068	0.3689	0.2661	0.000*	0.1650	0.3301	0.0431	0.001*	0.1165	6.2621	0.8768	0.000*
	DNT	-	5.9709	-	0.000*	-	0.3786	-	0.000*	-	0.2524	-	0.002*	-	6.4660	-	0.000*
Healthcare	NONE	0.7525	4.5941	0.0462	0.000*	0.0891	0.3366	0.2589	0.000*	0.0198	0.2376	0.7407	0.004*	0.8416	0.8416	0.0704	0.000*
	DNT	-	3.8317	-	0.000*	-	0.2772	-	0.000*	-	0.2277	-	0.001*	-	4.2376	-	0.000*
Business	NONE	1.2804	3.3364	0.0342	0.000*	0.1776	0.2897	0.0051	0.054	0.1308	0.2430	0.1226	0.119	1.6262	3.8224	0.0255	0.000*
	DNT	-	2.2897	-	0.000*	-	0.1682	-	0.205	-	0.1869	-	0.209	-	2.6075	-	0.000*
Adult	NONE	0.0500	0.9500	0.9239	0.018	0.0800	0.3400	0.6829	0.095	0.0200	0.1500	0.8861	0.309	0.2000	1.8900	0.7774	0.002*
	DNT	-	0.9000	-	0.025	-	0.2600	-	0.162	-	0.1300	-	0.314	-	1.6900	-	0.005*

* Significant at 0.01

5.3 Discussion of the Validation Study

Website Categories and Industry Sectors: The T/A sector is clearly dominant in terms of number of third-parties involved. One reason behind this is that the information involved in T/A sector is perhaps less sensitive than the F and P sectors. Publishers tend to stick with smaller number of third-parties in the F and P sectors. It is safe to assume that privacy concerns play an important role in the sharing behavior within each sector. Another reason could be that more money is potentially available for T/A than F and P. There are also big differences between publisher website subject categories. There is overall less sharing in adult, health, and business subject categories in the T/A sector. It seems that subject categories with sensitive visitor information tend to share less. News websites on the other hand, share more than the other categories. Note that news agencies have been forced into revenue models involving extensive sharing in order to survive. The T/A sector is less concentrated than the other sectors, and concentrations vary by subject category. This can be attributed to varying information sensitivity and subject-matter specialization.

Effect of Privacy Tools: Adblock Plus does substantially reduce the amount of third-party sharing for T/A sector. DNT, by and large, has no reducing effect on third-party sharing. Given that the maximum amount of third-party sharing actually increases substantially under the DNT condition, we conclude that in some specific cases the DNT flag can have a perverse effect and actually increase third-party sharing at some particular publisher websites. Policy implications are that optional use by publisher websites of passive privacy tools cannot be relied upon to reduce third-party sharing.

6. Discussion and Conclusion

We present a two-sided economic model to explain and analyze the decision-making of publisher websites who must balance user pricing and privacy. On the one hand, the publisher website needs to maintain their user base to increase profits. They must balance this with monetization through third-party information sharing and the subsequent personal privacy violations that result from this sharing, along with the associated declines in the user based due to third-party monetization, on the other hand. The proposed model is able to explain the underlying factors that affect how publisher websites use third-parties. The model describes how privacy concerns drive third-party market structures, with higher privacy concerns driving higher industry concentration. We have also provided several policy and social welfare implications and analyze the effect of public policy decisions such as taxation on the various stakeholders.

This paper makes several important contributions to our understanding of this problem. The two-sided economic model is created which allows researchers, managers, and policy makers to understand the impact of privacy violations by third-party vendor management by publisher websites. We show how increased users' privacy concerns decrease the number of vendors utilized by publisher websites, and how this in turn can lead to substantially higher industry concentration. We validate key outcomes of the model with an empirical validation study which confirms these key model outcomes regarding users' intent to disclose, third-party utilization, and industry concentration.

There exists a natural braking system to the actual privacy violation allowed by the publisher website. We might expect that higher per unit revenue earned by third-parties would actually encourage more violation. However, the counterintuitive finding is that higher per unit violation revenues for third-parties will have a braking effect on actual violation by reducing the violation at an increasing rate. When users perceive higher privacy violations, publisher website reduce the amount of violation that third-parties inflict upon users by reducing third-party sharing. Additionally, the equilibrium amount of the actual user privacy violation decreases as the users' perceived privacy violation from each third-party increases. With actual privacy violation levels being capped by consumer reaction, the equilibrium price charged to users by the publisher website increases with per user website operating costs.

There are limitations to this research. Note that different business models may dominate one website category versus another. The concern is that differences in third-party sharing may be due to differing business models, rather than users' privacy concerns related to the nature of the website subject. We recognize this limitation in the exploratory validation study. The impact of business model design on third-party utilization is beyond the scope of this study, and thus our economic model is limited to the impact of privacy concerns on third-party information sharing. Additionally, this study only examines direct information sharing. Consumer data is sold and resold, potentially making the problem much worse than is modeled in this paper. Because we do not consider the interaction effect between third-parties, our paper represents a lower bound on the problem, with the very real likelihood that results found with the model are understated.

We examine the impact of government taxation policy, and find that the impacts of a sales tax on the activity between the user and website differ from a tax on the third-party activity. Note that the effect of taxation is similar to putting a standard (that causes stakeholders some money) and risk reserves (which again causes the stakeholders some money). For policy makers, a key message is that postponing decisions to place limits on privacy violations caused by increasingly intrusive big data analytics technologies will result in increasingly higher price shocks to users. Profit and social welfare impacts of two intervention strategies are examined

The model presented in this paper predicts that user privacy concerns will affect third-party utilization, but is this truly validated by observed data? Actual sharing behaviour was observed for 700 websites - the top 100 in each of seven significant subject categories (news, arts, shopping, kids and teens, health, business, and adult) with 1066 unique third-party owner companies included, broken into three primary industry sectors of T/A (targeting/advertising), F (functionality), and P (performance). In the exploratory empirical validation of the model, we find that users' information is being shared extensively among third-parties by publisher websites, and the actual sharing behaviour is consistent with the model and predictions derived from the model. Due to privacy concerns and re-identification potential, the extent of third-party sharing is of strong interest to policy makers and regulatory organizations. It is also of foremost importance for website administrators. Sharing generates ad revenue and potentially better service, however, too much sharing will violate user privacy and reduces usage. We see these market forces being reflected in sharing levels and industry concentration measures between internet industries and content subject categories.

It is shown that website concerns for the users' reactions to privacy concerns affect the publisher website's willingness to share user information with third-party partners. Not only is the actual privacy violation important, but the perceived violation is also important. Industry structure is also affected by user privacy concerns. As measured by the HHI measurement, industry concentration increases as privacy concerns increase.

Future avenues for research include the following ideas. First, incorporate business revenue models into the economic model. Business models will likely have an impact on third-party utilization and industry concentration. Second, explore the issues surrounding passive privacy methods such as "Do Not Track" versus active privacy methods such as Adblock Plus. Extend the empirical study to determine sharing behaviour of specific websites with third-party partners, especially as it relates to different business models. Third, examine the security risks to the publisher website and users that result from the use of third-party vendors across the three different industry types.

References:

Alonso-Zaldivar, R, and Gillum, J. "Government health care website quietly sharing personal data" *Associated Press*, January 20, 2015. Accessessed 2015/01/26 online at http://abcnews.go.com/Technology/wireStory/privacy-concerns-governments-health-care-website-28340119

Anderson Jr, E. G., Parker, G. G., & Tan, B. (2013). Platform performance investment in the presence of network externalities. *Information Systems Research*, *25*(1), 152-172.

Chen, Jianqing; Jan Stallaert (2014). "An Economic Analysis of Online Advertising Using Behavioral Targeting". *MIS Quarterly* 38 (2): 429–449.

Krishnamurthy, B., & Wills, C. E. (2006). Generating a privacy footprint on the Internet. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement* (pp. 65-70). ACM.

Krishnamurthy, B., Malandrino, D., & Wills, C. E. (2007). Measuring privacy loss and the impact of privacy protection in web browsing. In*Proceedings of the 3rd symposium on Usable privacy and security* (pp. 52-63). ACM.

Krishnamurthy, B., & Wills, C. E. (2009a). On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on Online social networks* (pp. 7-12). ACM.

Krishnamurthy, B., & Wills, C. (2009b). Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th international conference on World wide web* (pp. 541-550). ACM.

Krishnamurthy, B., Naryshkin, K., & Wills, C. (2011, May). Privacy leakage vs. protection measures: the growing disconnect. In *Proceedings of the Web* (Vol. 2, pp. 1-10).

Li, Y. (2012). Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, *54*(1), 471-481.

Leon, P., Ur, B., Shay, R., Wang, Y., Balebako, R., & Cranor, L. (2012). Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 589-598). ACM.

Malandrino, D., & Scarano, V. (2013). Privacy leakage on the Web: Diffusion and countermeasures. *Computer Networks*, *57*(14), 2833-2855.

Mayer, J. R., & Mitchell, J. C. (2012). Third-party web tracking: Policy and technology. In *Security and Privacy (SP), 2012 IEEE Symposium on* (pp. 413-427). IEEE.

McDonald, A. M., & Cranor, L. F. (2010, October). Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *Proceedings of the 2010 Research Conference on Communication, Information and Internet Policy*.

Mikians, J., Gyarmati, L., Erramilli, V., & Laoutaris, N. (2012, October). Detecting price and search discrimination on the Internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks* (pp. 79-84). ACM.

Mueller, W. F., & Hamm, L. G. (1974). Trends in industrial market concentration, 1947 to 1970. *The Review of Economics and Statistics*, 511-520.

Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, *51*(10), 1494-1504.

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, *1*(4), 990-1029.

Schoen, Seth (2009). New Cookie Technologies: Harder to See and Remove, Widely Used to Track You. Electronic Frontier Foundation. Accessed online on 2014/7/10 at https://www.eff.org/deeplinks/2009/09/new-cookie-technologies-harder-see-and-remove-wide

Smith, Aaron (2014). Older adults and technology use. *Pew research internet project.*

Turow, J., King, J., Hoofnagle, C. J., Bleakley, A., & Hennessy, M. (2009). Americans reject tailored advertising and three activities that enable it. *Available at SSRN 1478214*.

Valentino-DeVries, J, Singer-Vine, J, and Soltani, A. "Websites Vary Prices, Deals Based on Users' Information." *Wall Street Journal*, December 24, 2012, sec.Tech. http://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

Appendices.

Appendix 1: Proof for Lemma 1.

The optimal violation factor V and publisher website price P_W satisfy the first order conditions:

$$\frac{\partial \Pi_W}{\partial V}(V^*, P_W^*) = 0 \ , \ \frac{\partial \Pi_W}{\partial P_W}(V^*, P_W^*) = 0$$

By simultaneously solving the two equations, V^* and P_W^* are calculated as given in Lemma 1. To ensure optimality, the second order conditions must hold:

$$\frac{\partial^2 \Pi_W}{\partial V^2} = \frac{M_D M_U r\,(v\,(P_W - C_W) + R_V (X - P_W))}{M_D M_U R_V\, v\, V^2 + F\, X}$$

$$- \frac{2\, M_D M_U r\ R_V\, v\, V (F\, M_U (P_W - C_W)(X - P_W) + M_D M_U r\ V (R_V (X - P_W) + v (P_W - C_W)) - M_D\, X\, r^2)}{(M_D M_U R_V\, v\, V^2 + F\, X)^2} < 0 \qquad \text{(A.1)}$$

$$\frac{\partial^2 \Pi_W}{\partial P_W^2} = - \frac{2\, F\, M_U}{M_D M_U R_V\, v\, V^2 + F\, X} < 0 \qquad \text{(A.2)}$$

$$det(Hessian) = \frac{\partial^2 \Pi_W}{\partial V^2} \frac{\partial^2 \Pi_W}{\partial P_W^2} - \left(\frac{\partial^2 \Pi_W}{\partial P_W \partial V}\right)^2 = - \frac{M_D\, M_U^{\,2}}{(M_D M_U R_V\, v\, V^2 + F\, X)^4}$$

$$(- M_D (2\, F\, M_U\, R_V\, v\, V\ (X + C_W - 2 P_W) + (R_V - v)(F r X - M_D M_U\, r\, R_V v\, V^2))^2 +$$

$$4\, F\, R_V\, v\, (3 X M_D^2 M_U\, r^2\, R_V\, v\, V^2 - F M_D\, r^2\, X^2 + F\, M_U\, (C_W - P_W)(P_W - X)(F X - 3 M_U M_D\, R_V v\, V^2) +$$

$$(P_W (R_V - v) + C_W v - R_V X)(M_D^{\,2} M_U^{\,2}\, r\, R_V v\, V^3 - 3 F M_D M_U\, r\, V\, X))) \geq 0 \qquad \text{(A.3)}$$

We also need the optimal number of users $N_U^{\,*} = N_U(V^*, P_W^*)$ and number of third-parties $N_D^{\,*} = N_D(V^*, P_W^*)$ to be positive. So we need to have:

$$N_U^* = \frac{M_U (X - C_W)}{2\, X} \geq 0 \Rightarrow X - C_W \geq 0 \qquad \text{(A.4)}$$

$$N_D^* = \frac{M_D\, r\, (R_V - v)}{2\, F\, v} \geq 0 \Rightarrow R_V - v \geq 0 \qquad \text{(A.5)}$$

So in section 3 we assume (A.1), (A.3), (A.4) , (A.5) to be true. (A.2) is always true.

◊