

When would online platforms pay data dividends?

Sukanya Kudva and Anil Aswani

Abstract—Online platforms, including social media and search platforms, have routinely used their users’ data for targeted ads, to improve their services, and to sell to third-party buyers. But an increasing awareness of the importance of users’ data privacy has led to new laws regulating platform data-sharing. Further, there have been political discussions on introducing data dividends, that pay users for their data. Three interesting questions are then: When would these online platforms be incentivized to pay data dividends? How does their decision depend on whether users value their privacy more than the platform’s free services? And should platforms invest in protecting users’ data? This paper considers various factors affecting the users’ and platform’s decisions through utility functions. We construct a principal-agent model using a Stackelberg game to calculate their optimal decisions and qualitatively discuss the implications. Our results could inform a policymaker trying to understand the consequences of mandating data dividends.

I. INTRODUCTION

The revenue models of many online platforms depend on collecting, analyzing, and selling users’ data. A free-service and advertising-based revenue model can cause conflicts in the users’ and platform’s interests. Users may be concerned about their privacy and possible misuse of data, while platforms want to maximize their profits. Further, users’ perception of a platform’s ethics and their willingness to participate can be affected by the platform’s revenue model, pricing decisions, and privacy practices [1]–[3].

A. Cybersecurity on online platforms

After the onset of the Internet of Things (IoT), there has been an increase in the variety, speed, and volume of users’ data collected. Information from multiple sources including devices, sensors, and social networks is being used by platforms to assist users and collect data [4], [5].

Though users have become more aware of the rampant collection of their data, they often do not know about the proliferation of IoT devices in their everyday lives. For instance, in August 2022 the Australian federal court convicted a major search platform for collecting users’ location data without their knowledge [6]. When users give consent and permissions to apps and platforms, they tend to underestimate the implications of it [17]. Reforms to protect users’ privacy are very much needed across the world [8]. Users should be able to control, delete and transfer their data across different platforms and service providers. They should

*This material is based upon work supported by the National Science Foundation under Grant CMMI-1847666.

S Kudva and A Aswani are with the Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, USA sukanya_kudva@berkeley.edu, aaswani@berkeley.edu

be asked for explicit consent every time a platform wants to use their data for a new purpose [4].

B. Privacy and data dividends

With growing user concerns, consumer privacy legislation has become an important topic for public discussion, and multiple new data privacy laws have been introduced [9]. In Europe, the General Data Protection Regulation (GDPR) was introduced in 2016 to give consumers more control over their data [18]. In 2019, legislators in California announced their intent to introduce data dividends, which is a model in which platforms would pay users in exchange for use of users’ data [11]. The same year, Oregon legislators introduced a bill, called the Health Information Property Act, to compensate consumers for monetizing their health data [12].

Implementing data dividends comes with its own challenges as companies holding the data are far more powerful than individual users. Further, there is a huge information asymmetry and only companies know the actual value of the users’ data. The critics of data dividends argue that selling data would make it a commodity and be counter-productive in protecting users’ privacy. They also feel that vulnerable groups – such as people of color and the poor – who are currently discriminated against should not be incentivized to pour more data into the system and further reduce their privacy [13], [14]. On the other hand, the proponents of data dividends argue that today’s technology economy is hugely driven by monetizing users’ data, and paying users a share of these benefits is only fair.

Recent studies have explored different ways of pricing data dividends for each user based on the value of their individual data [15]–[18]. Using the idea of Shapley and Owen values, they calculated the contribution made by each user to the platform’s profits. For our work, we ask different questions: Should platforms pay data dividends at all, and why? In this paper, we do our analysis with homogeneous users but our methods can be extended to analyze heterogeneous users too.

C. Contributions and outline

Our paper is organized as follows: Sect. II outlines our utility functions for an online platform and its users and our principal-agent model, Sect. III analytically solves the principal-agent model in order to derive their optimal choices and Sect. IV discusses insights from our model.

Our paper comes in the context of rising debates on data dividends. We try to understand when and how much online platforms would pay as data dividends. We consider users’ privacy concerns as an important factor, for which a platform can invest in data protection. The platform can also pay users

to incentivize taking risks and sharing their data. Our main contributions are to introduce a principal-agent model using a Stackelberg game to capture the platform-users dynamics and then use this to gain a better understanding of when the platform would pay users with data dividends.

II. OUR MODEL

An online platform provides its users with free services and a data dividend in exchange for their data. The users are allowed to choose between two levels – high or low – of data sharing. For each level, the platform provides a different data dividend and set of services. As the platform collects users' data, it faces the risk of a possible data breach. If a data breach occurs, the platform loses its reputation and faces possible legal and financial complications, and the users are harmed by the misuse of their data. Hence, the platforms consider investing in protecting their users' data. This could include the costs of building better technological infrastructure and signing insurance contracts.

A. Platform's utility

The platform has a fixed cost of S to maintain and provide its services. It invests I in users' data protection, reducing the probability of a data breach to $B(I)$. Here, B is assumed to be a twice differentiable function such that $B(I) > 0$ with $\lim_{I \rightarrow \infty} B(I) = 0$, $B'(I) < 0$ and $B''(I) > 0 \forall I$. Note that B is not a complementary cumulative distribution function. Instead, $B(I)$ gives the probability of data breach on investment I . In case of a data breach, the platform has a loss of F from legal cases and a lost reputation. When having access to k users' data, the platform makes a revenue of $U(k, b)$, where b of the total k users chose the low level of data sharing. The revenue may be due to selling the data to a third-party, selling advertisements to display to users or other sources. The platform pays a share of this revenue to users as data dividends, which are priced at two scales – p_0 and p_1 – for the low and high levels of data sharing respectively.

Considering these costs, revenue, and risks, the platform has a total expected utility of :

$$U(k, b) - B(I)F - I - S - p_0b - p_1(k - b). \quad (1)$$

B. User's utility

We consider k homogeneous users who have the same behavior and parameters. Each user values their personal data at V and faces an additional personal loss of L on a data breach. They also benefit from the platform's free services, which amount to a value of W . When a user chooses the low level of data sharing, these benefits and losses scale down by a factor $\alpha \in (0, 1)$. The platform pays users with data dividends at two scales of pay and thereby encourages a particular level of data sharing.

Let c_i be user i 's decision variable so that $c_i = 0$ and 1 for the low and high levels of data sharing respectively. Then a user i 's total expected utility is:

$$c_i(p_1 - \mathcal{V}) + (1 - c_i)(p_0 - \alpha\mathcal{V}), \quad (2)$$

where $\mathcal{V} = \bar{V} + B(I)L$ and $\bar{V} = V - W$.

C. Principal-agent formulation

We construct a Stackelberg game [19]–[21], in which players act sequentially with follower(s) acting after a leader. In our model, the platform first prices data dividends for different levels of data-sharing, and decides on investment for users' data protection. Given this information, the users decide how much data to share.

The optimal, equilibrium choices of the platform and the users are their Stackelberg strategies. We capture this using an optimization problem, which maximizes the platform's utility when the users are maximizing their utilities [22]. It is formulated as follows:

$$\begin{aligned} \max \quad & U(k, b) - B(I)F - I - S - p_0b - p_1(k - b) \\ \text{s.t.} \quad & c_i^* = \arg \max_{c_i \in \{0, 1\}} c_i(p_1 - \mathcal{V}) + (1 - c_i)(p_0 - \alpha\mathcal{V}) \\ & \forall i = \{1, \dots, k\} \\ & c_i^*(p_1 - \mathcal{V}) + (1 - c_i^*)(p_0 - \alpha\mathcal{V}) \geq 0 \\ & \forall i = \{1, \dots, k\} \\ & I, p_0, p_1 \geq 0. \end{aligned} \quad (3)$$

The second constraint in (3) ensures that the users do not have a net loss from using the platform, without which they would leave the platform.

III. OPTIMAL CHOICES

Since the users are homogeneous, they make similar choices: either $c_i^* = 1$ or $c_i^* = 0$ for every user i . If users find both levels of data sharing to be utility-maximizing, then we assume that they all choose the level that most benefits the platform. We solve these two cases of c_i^* separately. The optimal solution is the best of the two cases and can vary with the numerical values of the parameters of the model.

A. Case 1: $c_i^* = 1 \forall i = \{1, \dots, k\}$

To ensure users choose high level of data sharing, the platform must price the data dividends such that $p_1 - \mathcal{V} \geq p_0 - \alpha\mathcal{V}$. Note that $b = 0$ in this case. The optimization problem then reduces to:

$$\begin{aligned} \max \quad & U(k, 0) - B(I)F - I - p_1k - S \\ \text{s.t.} \quad & p_1 - \mathcal{V} \geq p_0 - \alpha\mathcal{V} \\ & p_1 - \mathcal{V} \geq 0 \\ & I, p_0, p_1 \geq 0. \end{aligned} \quad (4)$$

Since $p_0 \geq 0$, $p_0^* = 0$ is optimal for the problem. Given this, one can conclude $p_1^* = \mathcal{V}$ when $\mathcal{V} \geq 0$ and $p_1^* = 0$ when $\mathcal{V} \leq 0$. Also, $U(k, 0) - S$ can be treated as a constant. These observations further reduce (4) to two sub-cases with optimization problems in a single variable I .

1) Sub-case 1 - If $\mathcal{V} \geq 0$ then $p_1^* = \mathcal{V}$: Set $p_1^* = \mathcal{V} = \bar{V} + B(I)L$ and add an additional constraint $\mathcal{V} \geq 0$ in (4).

$$\begin{aligned} \min \quad & B(I)F_1 + I + \bar{V}k \\ \text{s.t.} \quad & \bar{V} + B(I)L \geq 0 \\ & I \geq 0, \end{aligned} \quad (5)$$

where $F_1 = F + Lk$.

Let $I_1 = B'^{-1}(\frac{-1}{F_1})$ and $I_2 = B^{-1}(\frac{-\bar{V}}{L})$.

Proposition III.1. When $\bar{V} \geq 0$, $I^* = I_1$ if I_1 exists. Else, $I^* = 0$.

Proof. $B(I)L \geq -\bar{V}$ holds $\forall I$ because $\bar{V} \geq 0$. When I_1 exists, it's the optimal solution to the unconstrained problem. Since it also satisfies $I_1 \geq 0$, it's the optimal solution for the constrained problem too. When I_1 doesn't exist, the objective in (5) is an increasing function for $I \geq 0$. So the smallest feasible value of I , i.e 0, is optimal. \square

Proposition III.2. When $\bar{V} \leq 0$,

- 1) If I_1 exists, $I^* = I_1$ when $B(I_1)L \geq -\bar{V}$. Else when $B(I_1)L < -\bar{V}$, $I^* = I_2$ if I_2 exists and no solution otherwise.
- 2) If I_1 doesn't exist, $I^* = 0$ when $B(0)L \geq -\bar{V}$ and no solution otherwise.

Proof. When I_1 exists and $B(I_1)L \geq -\bar{V}$ holds, I_1 , which is the unconstrained optimal solution, satisfies both constraints and is optimal. When $B(I_1)L < -\bar{V}$, $I \geq I_1$ doesn't satisfy the first constraint in (5) and the objective is a decreasing function for $I < I_1$. If I_2 exists, it is the largest $I < I_1$ satisfying $\bar{V} + B(I_2)L \geq 0$ and is optimal. If I_2 , which satisfies $\bar{V} + B(I_2)L = 0$, doesn't exist then there is no solution as $B(I)$ is a strictly decreasing function and the constraint $B(I)L \geq -\bar{V}$ can never be satisfied.

When I_1 doesn't exist, the objective is an increasing function for $I \geq 0$. If $B(0)L \geq -\bar{V}$ then $I = 0$ satisfies both constraints of the problem and is optimal. If $B(0)L < -\bar{V}$ then there is no solution as $B(I)$ is strictly decreasing and the constraint $B(I)L \geq -\bar{V}$ can never hold. \square

2) Sub-case 2 - If $\mathcal{V} \leq 0$ then $p_1^* = 0$: Set $p_1 = 0$ and add an additional constraint $\mathcal{V} \leq 0$ in (4).

$$\begin{aligned} & \min B(I)F_0 + I \\ & \text{s.t. } \bar{V} + B(I)L \leq 0 \\ & \quad I \geq 0 \end{aligned} \quad (6)$$

where $F_0 = F$.

Let $I_3 = B'^{-1}(\frac{-1}{F_0})$ and as before, $I_2 = B^{-1}(\frac{-\bar{V}}{L})$.

Proposition III.3. When $\bar{V} > 0$, there is no solution.

Proof. $-\bar{V} \geq B(I)L$ can never hold for any I as $\bar{V} > 0$. \square

Proposition III.4. When $\bar{V} \leq 0$,

- 1) If I_3 exists, $I^* = I_3$ when $-\bar{V} \geq B(I_3)L$. Else when $-\bar{V} < B(I_3)L$, $I^* = I_2$.
- 2) If I_3 doesn't exist, $I^* = 0$ when $-\bar{V} \geq B(0)L$ and $I^* = I_2$ otherwise.

Proof. When I_3 exists and $-\bar{V} \geq B(I_3)L$ holds, I_3 , which is the unconstrained optimal solution, satisfies both constraints and is optimal. When $-\bar{V} < B(I_3)L$, $I \leq I_3$ doesn't satisfy $-\bar{V} \geq B(I)L$ as $B(I)$ is strictly decreasing. Since the objective is increasing for $I > I_3$, I_2 which is the smallest I satisfying $-\bar{V} \geq B(I)L$ is optimal. Here, I_2 exists because $I_2 > I_3$ and $B(I)$ is strictly decreasing.

When I_3 doesn't exist, the objective is increasing for $I \geq 0$ and hence the smallest feasible I is optimal. If $-\bar{V} \geq B(0)L$ then $I = 0$ satisfies both the constraints and is optimal. If $-\bar{V} < B(0)L$ then $I = I_2$ is the smallest feasible I with $-\bar{V} + B(I)L = 0$ and is optimal. \square

The results from the two sub-cases can be combined by finding the minimum objective value of the two sub-cases.

3) Final results of Case 1:

Theorem III.5. For the case when all users choose the high level of data-sharing, when $\bar{V} \leq 0$ the optimal choices of the platform are as follows:

- 1) If I_1 and I_3 exist, $(I^* = I_1, p_1^* = \bar{V} + B(I_1)L, p_0^* = 0)$ when $L \geq \frac{-\bar{V}}{B(I_1)}$, $(I^* = I_2, p_1^* = 0, p_0^* = 0)$ when $\frac{-\bar{V}}{B(I_1)} \geq L \geq \frac{-\bar{V}}{B(I_3)}$ and $(I^* = I_3, p_1^* = 0, p_0^* = 0)$ when $\frac{-\bar{V}}{B(I_3)} \geq L$.
- 2) If I_1 exists but not I_3 , $(I^* = I_1, p_1^* = \bar{V} + B(I_1)L, p_0^* = 0)$ when $L \geq \frac{-\bar{V}}{B(I_1)}$, $(I^* = I_2, p_1^* = 0, p_0^* = 0)$ when $\frac{-\bar{V}}{B(I_1)} \geq L \geq \frac{-\bar{V}}{B(0)}$ and $(I^* = 0, p_1^* = 0, p_0^* = 0)$ when $\frac{-\bar{V}}{B(0)} \geq L$.
- 3) If both I_1 and I_3 don't exist, $(I^* = 0, p_1^* = \bar{V} + B(0)L, p_0^* = 0)$ when $L \geq \frac{-\bar{V}}{B(0)}$ and $(I^* = 0, p_1^* = 0, p_0^* = 0)$ when $\frac{-\bar{V}}{B(0)} \geq L$.

Proof. The results from Propositions III.1-III.4 are combined by finding the optimal of the two sub-cases.

1) When I_1 and I_3 exist :

If $L \geq \frac{-\bar{V}}{B(I_1)}$, the optimal solution is one of $(I = I_1, p_1 = \bar{V} + B(I_1)L, p_0 = 0)$ and $(I = I_2, p_1 = 0, p_0 = 0)$ with objective function values $B(I_1)F_1 + I_1 + \bar{V}k$ and $B(I_2)F_0 + I_2$ respectively. The former solution is optimal if:

$$\begin{aligned} B(I_2)F_0 + I_2 & \geq B(I_1)F_1 + I_1 + \bar{V}k \\ \iff I_2 - I_1 & \geq F_1(B(I_1) - B(I_2)) \\ \iff \frac{-1}{F_1} & \leq \frac{B(I_1) - B(I_2)}{I_1 - I_2} \end{aligned} \quad (7)$$

Since $B(I_1) \geq \frac{-\bar{V}}{L} = B(I_2)$, $I_1 \leq I_2$ as B is strictly decreasing. Using the mean value theorem, $\exists I' \text{ s.t } I_1 \leq I' \leq I_2$ and $B'(I') = \frac{B(I_1) - B(I_2)}{I_1 - I_2}$. As $B'' > 0$, $B'(I_1) \geq B'(I') \implies \frac{-1}{F_1} \leq \frac{B(I_1) - B(I_2)}{I_1 - I_2}$. Hence, (7) is true.

If $\frac{-\bar{V}}{B(I_1)} \geq L \geq \frac{-\bar{V}}{B(I_3)}$ then both sub-cases give the same optimal solution $(I^* = I_2, p_1^* = 0, p_0^* = 0)$.

If $\frac{-\bar{V}}{B(I_3)} \geq L \geq \frac{-\bar{V}}{B(0)}$ then the optimal solution is one of $(I = I_3, p_1 = 0, p_0 = 0)$ and $(I = I_2, p_1 = 0, p_0 = 0)$ with objective function values $B(I_3)F_0 + I_3$ and $B(I_2)F_0 + I_2$ respectively. The former solution is optimal if:

$$\begin{aligned} B(I_2)F_0 + I_2 & \geq B(I_3)F_0 + I_3 \\ \iff \frac{-1}{F_0} & \geq \frac{B(I_3) - B(I_2)}{I_3 - I_2} \end{aligned} \quad (8)$$

Since $B(I_3) \leq \frac{-\bar{V}}{L} = B(I_2)$, $I_3 \geq I_2$ as B is strictly decreasing. Using the mean value theorem, $\exists I' \text{ s.t } I_2 \leq I' \leq I_3$ and $B'(I') = \frac{B(I_3) - B(I_2)}{I_3 - I_2}$. As $B'' > 0$, $B'(I_3) \geq B'(I') \implies \frac{-1}{F_0} \geq \frac{B(I_3) - B(I_2)}{I_3 - I_2}$. Hence, (8) is true.

If $\frac{-\bar{V}}{B(0)} \geq L$ then $(I^* = I_3, p_1^* = 0, p_0^* = 0)$ is only solution from the two sub-cases as one of them was infeasible.

2) When I_1 exists but not I_3 :

If $L \geq \frac{-\bar{V}}{B(I_1)}$ then $(I = I_1, p_1 = \bar{V} + B(I_1)L, p_0 = 0)$ and $(I = I_2, p_1 = 0, p_0 = 0)$ are compared and (7) holds. If $\frac{-\bar{V}}{B(I_1)} \geq L \geq \frac{-\bar{V}}{B(0)}$ then both sub-cases give the same optimal solution $(I^* = I_2, p_1^* = 0, p_0^* = 0)$. If $\frac{-\bar{V}}{B(0)} \geq L$ then $(I^* = 0, p_1^* = 0, p_0^* = 0)$ is only solution from the two sub-cases as one of them was infeasible.

3) When both I_1 and I_3 don't exist:

If $L \geq \frac{-\bar{V}}{B(0)}$ then $(I = 0, p_1 = \bar{V} + B(0)L, p_0 = 0)$ and $(I = I_2, p_1 = 0, p_0 = 0)$ with objective function values $B(0)F_1 + \bar{V}k$ and $B(I_2)F_0 + I_2$ respectively are compared. The former solution is optimal if:

$$\begin{aligned} & B(I_2)F_0 + I_2 \geq B(0)F_1 + \bar{V}k \\ \iff & (B(I_2) - B(0))F_1 \geq -I_2 \\ \iff & \frac{B(I_2) - B(0)}{I_2} \geq \frac{-1}{F_1} \end{aligned} \quad (9)$$

Using the mean value theorem, $\exists I' \text{ s.t } I_2 \leq I' \leq 0$ and $B'(I') = \frac{B(I_2) - B(0)}{I_2}$. And $B'(I') \geq \frac{-1}{F_1}$ as I_1 doesn't exist. Hence, (9) is true.

If $\frac{-\bar{V}}{B(0)} \geq L$ then $(I^* = 0, p_1^* = 0, p_0^* = 0)$ is only solution from the two sub-cases as one of them was infeasible. \square

Theorem III.6. For the case when all users choose the high level of data-sharing, when $\bar{V} \geq 0$ the optimal choices of the platform are as follows:

- 1) If I_1 exists then $(I^* = I_1, p_1^* = \bar{V} + B(I_1)L, p_0^* = 0)$ is optimal.
- 2) If I_1 doesn't exist then $(I^* = 0, p_1^* = \bar{V} + B(0)L, p_0^* = 0)$

Proof. The results follow from Propositions III.1-III.4, and the above solutions are the only feasible solutions from the two sub-cases. \square

The final results of case 1 are summarized in table I.

B. Case 2: $c_i^* = 0 \ \forall i = \{1, \dots, k\}$

In this case, platform must price the data dividends such that $p_0 - \alpha\mathcal{V} \geq p_1 - \mathcal{V}$. Here, $b = k$. The optimization problem becomes:

$$\begin{aligned} & \max U(k, k) - B(I)F - I - p_0k - S \\ \text{s.t. } & p_0 - \alpha\mathcal{V} \geq p_1 - \mathcal{V} \\ & p_0 - \alpha\mathcal{V} \geq 0 \\ & I, p_0, p_1 \geq 0. \end{aligned} \quad (10)$$

Similar to case 1, it can be observed that $p_1^* = 0$ is optimal and $p_0^* = \alpha\mathcal{V}$ when $\mathcal{V} \geq 0$ and $p_0^* = (\alpha - 1)\mathcal{V}$ when $\mathcal{V} \leq 0$.

Let $I_4 = B'^{-1}\left(\frac{-1}{F + \alpha Lk}\right)$, $I_5 = B'^{-1}\left(\frac{-1}{F + (\alpha - 1)Lk}\right)$ and as before, $I_2 = B^{-1}\left(\frac{-\bar{V}}{L}\right)$.

Theorem III.7. For the case when all users choose the low level of data-sharing, when $\bar{V} \leq 0$ the optimal choices of the platform are as follows:

- 1) If I_4 and I_5 exist, $(I^* = I_4, p_1^* = 0, p_0^* = \alpha(\bar{V} + B(I_4)L))$ when $L \geq \frac{-\bar{V}}{B(I_4)}$, $(I^* = I_2, p_1^* = 0, p_0^* = 0)$ when $\frac{-\bar{V}}{B(I_4)} \geq L \geq \frac{-\bar{V}}{B(I_5)}$ and $(I^* = I_5, p_1^* = 0, p_0^* = (\alpha - 1)(\bar{V} + B(I_5)L))$ when $\frac{-\bar{V}}{B(I_5)} \geq L$.
- 2) If I_4 exists but not I_5 , $(I^* = I_4, p_1^* = 0, p_0^* = \alpha(\bar{V} + B(I_4)L))$ when $L \geq \frac{-\bar{V}}{B(I_4)}$, $(I^* = I_2, p_1^* = 0, p_0^* = 0)$ when $\frac{-\bar{V}}{B(I_4)} \geq L \geq \frac{-\bar{V}}{B(0)}$ and $(I^* = 0, p_1^* = 0, p_0^* = (\alpha - 1)(\bar{V} + B(0)L))$ when $\frac{-\bar{V}}{B(0)} \geq L$.
- 3) If both I_4 and I_5 don't exist, $(I^* = 0, p_1^* = 0, p_0^* = \alpha(\bar{V} + B(0)L))$ when $L \geq \frac{-\bar{V}}{B(0)}$ and $(I^* = 0, p_1^* = 0, p_0^* = (\alpha - 1)(\bar{V} + B(0)L))$ when $\frac{-\bar{V}}{B(0)} \geq L$.

Proof. Propositions similar to III.1-III.4 can be derived for case 2. The proof is then similar to Theorem III.5. \square

Theorem III.8. For the case when all users choose the low level of data-sharing, when $\bar{V} \geq 0$ the optimal choices of the platform are as follows:

- 1) If I_4 exists then $(I^* = I_4, p_1^* = 0, p_0^* = \alpha(\bar{V} + B(I_4)L))$ is optimal.
- 2) If I_4 doesn't exist then $(I^* = 0, p_1^* = 0, p_0^* = \alpha(\bar{V} + B(0)L))$

Proof. The proof is similar to Theorem III.6. \square

The final results of case 2 are summarized in table I.

IV. INSIGHTS FROM OPTIMAL CHOICES

The platform uses data dividends to direct users to choose a particular level of data sharing. In general, the platform pays a higher data dividend for the high level of data sharing to compensate for users' larger risks. But it also earns a larger revenue when it collects more data. This trade-off between revenue and data dividend decides the optimal level of sharing.

It is optimal for the platform to pay for only one level of data sharing which maximizes its profit. However, sometimes it may prefer to not pay at all. The cases when this happens are outlined below.

A. Can $p_1^* = 0$ while encouraging high level of data sharing?

Firstly, $p_1^* \neq 0$ when $V > W$. That is, the platform always pays a data dividend whenever users value their data more than the platform's free services. When $V \leq W$, the users value the platform's free services more than their data. In this case, if the platform ensures that a data breach does not occur then it does not pay any data dividend. So if $B(I) \rightarrow 0 \ \forall I$, meaning there is a low possibility of a data breach, or $L \rightarrow 0$ so that users face a small loss from a data breach, then no data dividend is paid. Interestingly, even if $F \rightarrow \infty$, which is if the platform has heavy losses from a data breach, then they do not pay users with data dividends and instead invest heavily in data protection ($I \rightarrow \infty$). In general, the platform does not pay a data dividend when $W - V - B(I^*)L \geq 0$, that is, the net users' utility before pay is positive. Currently, major technology companies do not pay data dividends. Our

TABLE I

Final results of Case 1 when $V \leq 0$: Here, $p_0^* = 0$ always.		
I_1 and I_3 exist	I_1 exists but not I_3	I_1 and I_3 don't exist
If $L \geq \frac{-V}{B(I_1)} : I^* = I_1$, $p_1^* = \bar{V} + B(I_1)L$.	If $L \geq \frac{-V}{B(I_1)} : I^* = I_1$, $p_1^* = \bar{V} + B(I_1)L$.	If $L \geq \frac{-V}{B(0)} : I^* = 0$, $p_1^* = \bar{V} + B(0)L$.
If $\frac{-\bar{V}}{B(I_1)} \geq L \geq \frac{-\bar{V}}{B(I_3)} : I^* = I_2$, $p_1^* = 0$.	If $\frac{-\bar{V}}{B(I_1)} \geq L \geq \frac{-\bar{V}}{B(0)} : I^* = I_2$, $p_1^* = 0$.	If $\frac{-\bar{V}}{B(0)} \geq L : I^* = 0$, $p_1^* = 0$.
If $\frac{-\bar{V}}{B(I_3)} \geq L : I^* = I_3$, $p_1^* = 0$.	If $\frac{-\bar{V}}{B(0)} \geq L : I^* = 0$, $p_1^* = 0$.	
Final results of Case 1 when $V \geq 0$: Here, $p_0^* = 0$ always.		
I_1 exists $I^* = I_1$ $p_1^* = \bar{V} + B(I_1)L$		I_1 doesn't exist $I^* = 0$ $p_1^* = \bar{V} + B(0)L$
Final results of Case 2 when $V \leq 0$: Here, $p_1^* = 0$ always.		
I_4 and I_5 exist	I_4 exists but not I_5	I_4 and I_5 don't exist
If $L \geq \frac{-V}{B(I_4)} : I^* = I_4$, $p_0^* = \alpha(\bar{V} + B(I_4)L)$.	If $L \geq \frac{-V}{B(I_4)} : I^* = I_4$, $p_0^* = \alpha(\bar{V} + B(I_4)L)$.	If $L \geq \frac{-\bar{V}}{B(0)} : I^* = 0$, $p_0^* = \alpha(\bar{V} + B(0)L)$.
If $\frac{-\bar{V}}{B(I_4)} \geq L \geq \frac{-\bar{V}}{B(I_5)} : I^* = I_2$, $p_0^* = 0$.	If $\frac{-\bar{V}}{B(I_4)} \geq L \geq \frac{-\bar{V}}{B(0)} : I^* = I_2$, $p_0^* = 0$.	If $\frac{-\bar{V}}{B(0)} \geq L : I^* = 0$, $p_0^* = (\alpha - 1)(\bar{V} + B(0)L)$.
If $\frac{-\bar{V}}{B(I_5)} \geq L : I^* = I_5$, $p_0^* = (\alpha - 1)(\bar{V} + B(I_5)L)$.	If $\frac{-\bar{V}}{B(0)} \geq L : I^* = 0$, $p_0^* = (\alpha - 1)(\bar{V} + B(0)L)$.	
Final results of Case 2 when $V \geq 0$: Here, $p_1^* = 0$ always.		
I_4 exists $I^* = I_4$ $p_0^* = \alpha(\bar{V} + B(I_4)L)$		I_4 doesn't exist $I^* = 0$ $p_0^* = \alpha(\bar{V} + B(0)L)$

model suggests that these companies believe they are in the region described by the above discussion.

B. Can $p_0^* = 0$ while encouraging low level of data sharing?

Like for p_1 , if $V > W$ then $p_0^* \neq 0$. But when $V \geq W$, $p_0^* = 0$ only if $W - V - B(I^*)L = 0$, that is, the net users' utility is zero. This is strange because the platform pays a data dividend when $W - V - B(I^*)L > 0$, that is, the net users' utility before pay is positive. The reason is that in this case the users naturally prefer a high level of data sharing. So the platform has to pay a data dividend to incentivize the low level of data sharing. One could argue that the platform should rather remove the option of a high level of data sharing, and it probably would.

Finally, we also analyze the trends in the investment for users' data protection. If $B(I) \rightarrow 0 \forall I$, meaning the possibility of a data breach is small, or if $F, L \rightarrow 0$ so that no party loses from a data breach, then the platform does not invest in data protection. If $B'(I) \rightarrow 0 \forall I$, that is when the likelihood of a data breach is not very sensitive to investment, then the platform chooses to not invest at all.

V. CONCLUSION

We studied the dynamics between an online platform and its users when the latter are paid with data dividends.

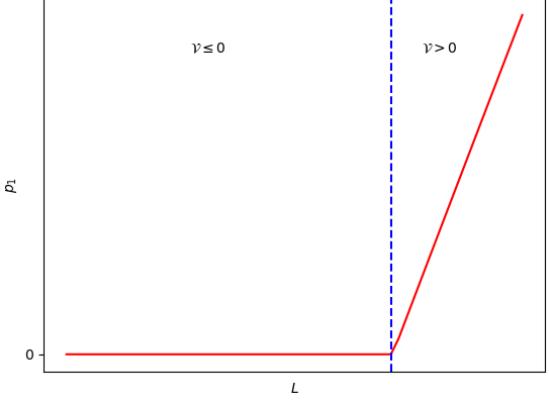


Fig. 1. Here, users choose the high level of data sharing, I_1 exists, and $\bar{V} \leq 0$. Data dividend p_1 is plotted for increasing values of L , that is users' loss from a data breach. The platform pays a data dividend beyond a threshold when $V > 0$, that is users face a total loss.

We demonstrated how an online platform could use data dividends to incentivize its users to share more data and take more risks, while also giving them an option to share as much data as they want. If the platform wants to gather more data, it pays a data dividend only when the users perceive their risks and losses to be higher than their benefits. Further, when

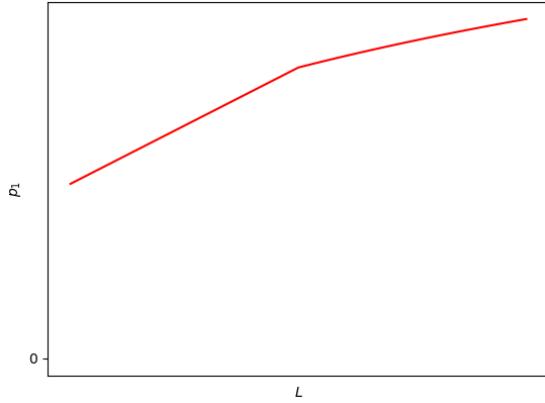


Fig. 2. Here, users choose the high level of data sharing and $\bar{V} \geq 0$. Data dividend p_1 is plotted for increasing values of L , that is users' loss from a data breach. The platform always pays a data dividend of at least \bar{V} .

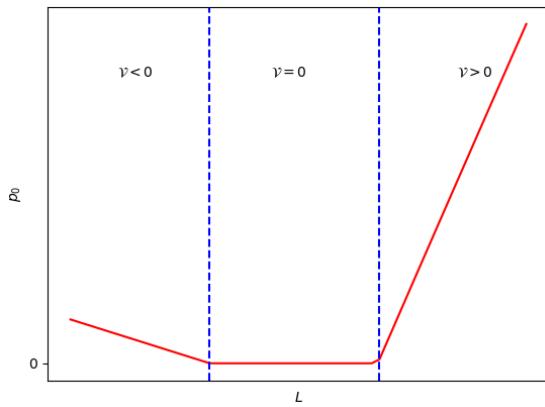


Fig. 3. Here, users choose the low level of data sharing, I_4 exists, and $\bar{V} \leq 0$. Data dividend p_0 is plotted for increasing values of L , that is users' loss from a data breach. Interestingly, the platform doesn't pay a data dividend only when $V = 0$, that is users neither have a net loss nor gain.

users feel that they do not lose much from a data breach, the platform prefers to invest in data protection rather than pay users for risking a breach. On the other hand, when users perceive a high loss from a breach, the platform has to both invest and pay. It would be interesting to analyze multiple (more than 2) levels of data sharing, each with a different data dividend, and where heterogeneous users are incentivized to share more or fewer data as is optimal for the platform. We leave this for future work.

It is debatable whether data dividends could treat data as a commodity and exploit the poor and vulnerable. But if data dividends are introduced, our model shows that platforms cannot always use it to shirk responsibility for data protection. When users feel considerable harm from a data breach, platforms would both invest in data protection and pay higher data dividends. We believe that our insights could help a policymaker understand data dividends better.

REFERENCES

- [1] L. Xia, K. B. Monroe, and J. L. Cox, "The Price is Unfair! A Conceptual Framework of Price Fairness Perceptions," *Journal of Marketing*, vol. 68, no. 4, pp. 1–15, 2004, DOI: 10.1509/jmkg.68.4.1.42733.
- [2] Y. Su and L. Jin, "The Impact of Online Platforms' Revenue Model on Consumers' Ethical Inferences," *J Bus Ethics*, vol. 178, no. 2, pp. 555–569, Jun. 2022, DOI: 10.1007/s10551-021-04798-0.
- [3] L. Yu, H. Li, W. He, F.-K. Wang, and S. Jiao, "A meta-analysis to explore privacy cognition and information disclosure of internet users," *International Journal of Information Management*, vol. 51, p. 102015, Apr. 2020, DOI: 10.1016/j.ijinfomgt.2019.09.011.
- [4] C. Perera, R. Ranjan, L. Wang, S. U. Khan, and A. Y. Zomaya, "Big Data Privacy in the Internet of Things Era," *IT Professional*, vol. 17, no. 3, pp. 32–39, May 2015, DOI: 10.1109/MITP.2015.34.
- [5] R. Dong, L. J. Ratliff, A. A. Cárdenas, H. Ohlsson, and S. S. Sastry, "Quantifying the Utility-Privacy Tradeoff in the Internet of Things," *ACM Trans. Cyber-Phys. Syst.*, vol. 2, no. 2, pp. 1–28, Apr. 2018, DOI: 10.1145/3185511.
- [6] A. A. Press, "Google to pay \$60m fine for misleading Australians about collecting location data," *The Guardian*, Aug. 12, 2022. Accessed: Sep. 24, 2022. [Online] Available: <https://www.theguardian.com/technology/2022/aug/12/google-to-pay-60m-fine-for-misleading-australians-about-collecting-location-data>.
- [7] J. Golbeck and M.L. Mauriello, "User Perception of Facebook App Data Access: A Comparison of Methods and Privacy Concerns", 2014, [Online] Available: <http://hcil2.cs.umd.edu/trs/2014-19/2014-19.pdf>.
- [8] L. Na, C. Yang, C.-C. Lo, F. Zhao, Y. Fukuoka, and A. Aswani, "Feasibility of reidentifying individuals in large national physical activity data sets from which protected health information has been removed with use of machine learning", *JAMA Network Open*, vol. 1, no. 8, pp. e186040–e186040, 2018.
- [9] "2022 Consumer Privacy Legislation." <https://www.ncsl.org/research/telecommunications-and-information-technology/2022-consumer-privacy-legislation.aspx>.
- [10] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016." <http://cnpd.public.lu/en/legislation/droit-europ/union-europeenne/rgrp.html>.
- [11] "Newsom wants companies collecting personal data to share the wealth with Californians," *Los Angeles Times*, May 05, 2019. <https://www.latimes.com/politics/la-pol-ca-gavin-newsom-california-data-dividend-20190505-story.html>.
- [12] H. Journal, "Oregon Health Information Property Act Proposes Paying Patients to Share Their Healthcare Data," *HIPAA Journal*, Jan. 31, 2019. <https://www.hipaajournal.com/oregon-health-information-property-act-proposes-paying-patients-to-share-their-healthcare-data/>.
- [13] H. Tsukayama, "Why Getting Paid for Your Data Is a Bad Deal," *Electronic Frontier Foundation*, Oct. 26, 2020. <https://www.eff.org/deeplinks/2020/10/why-getting-paid-your-data-bad-deal>.
- [14] R. Bartlett, A. Morse, R. Stanton, and N. Wallace, "Consumer-lending discrimination in the FinTech Era," *Journal of Financial Economics*, vol. 143, no. 1, pp. 30–56, Jan. 2022, DOI: 10.1016/j.jfineco.2021.05.047.
- [15] E. Bax, "Computing a Data Dividend," *arXiv*, Jun. 27, 2019. DOI: 10.48550/arXiv.1905.01805.
- [16] E. Bax et al., "Data Consortia," in *Proceedings of the Future Technologies Conference (FTC) 2020*, Volume 2, Cham, 2021, pp. 489–498. DOI: 10.1007/978–3–030–63089–8_31.
- [17] M. Farboodi, D. Singal, L. Veldkamp, and V. Venkateswaran, "Valuing Financial Data," *National Bureau of Economic Research*, Mar. 2022. DOI: 10.3386/w29894.
- [18] D. Acemoglu, A. Makhdoomi, A. Malekian, and A. Ozdaglar, "Too Much Data: Prices and Inefficiencies in Data Markets," *National Bureau of Economic Research*, Sep. 2019. DOI: 10.3386/w26296.
- [19] M. Simaan, "Equilibrium properties of the Nash and Stackelberg strategies," *Automatica*, vol. 13, no. 6, pp. 635–636, Nov. 1977, DOI: 10.1016/0005-1098(77)90086-3.
- [20] M. Simaan and J. B. Cruz, "On the Stackelberg strategy in nonzero-sum games," *J Optim Theory Appl*, vol. 11, no. 5, pp. 533–555, May 1973, DOI: 10.1007/BF00935665.
- [21] T. Basar, "On the relative leadership property of Stackelberg strategies," *J Optim Theory Appl*, vol. 11, no. 6, pp. 655–661, 1973, DOI: 10.1007/BF00935564.
- [22] Y. Lee and A. Aswani, "Optimally Designing Cybersecurity Insurance Contracts to Encourage the Sharing of Medical Data," in *Proceedings of the Conference on Decision and Control*, 2022. To appear.