Reputation and Data-Protection Incentives

Manos Perdikakis,
University of Oxford

EEA Conference, 2024

August 28th, 2024

Introduction

- Data breaches and cyber attacks are commonplace: (e.g. LastPass, Twitter, Facebook, Snapchat etc..)
- Consumers harmed by loss of personal information.
  - Difficult to ex ante identify cyber-secure firms.
  - but also tough for regulators to ex post verify how diligent a firm was.

Thus, **reputation** for respecting users' privacy becomes important.

- e.g. Facebook's "Privacy is Personal" campaign was about restoring *trust* after Cambridge Analytica.

Introduction

- Data breaches and cyber attacks are commonplace: (e.g. LastPass, Twitter, Facebook, Snapchat etc..)
- Consumers harmed by loss of personal information.
  - Difficult to ex ante identify cyber-secure firms.
  - but also tough for regulators to ex post verify how diligent a firm was.

Thus, **reputation** for respecting users' privacy becomes important.

- e.g. Facebook's "Privacy is Personal" campaign was about restoring *trust* after Cambridge Analytica.
- Consumers do observe whether **data breaches** occur: Their frequency affects firms' reputations.

Introduction

- Data breaches and cyber attacks are commonplace: (e.g. LastPass, Twitter, Facebook, Snapchat etc..)
- Consumers harmed by loss of personal information.
  - Difficult to ex ante identify cyber-secure firms.
  - but also tough for regulators to ex post verify how diligent a firm was.

Thus, **reputation** for respecting users' privacy becomes important.

- e.g. Facebook's "Privacy is Personal" campaign was about restoring *trust* after Cambridge Analytica.
- Consumers do observe whether **data breaches** occur: Their frequency affects firms' reputations.

This paper: develop a model of reputational concerns and evaluate GDPR-style policies around cyber security and data collection.

Motivation: impact of breaches on reputation

Question: Do firms that suffer cyber attacks suffer *reputational* damage?

Kamiya et al. 2021, JFinEcon: when a successful cyber attack involves loss of personal financial information, total shareholder loss is much larger than out-of-pocket costs.

Motivation: impact of breaches on reputation

Question: Do firms that suffer cyber attacks suffer *reputational* damage?

Kamiya et al. 2021, JFinEcon: when a successful cyber attack involves loss of personal financial information, total shareholder loss is much larger than out-of-pocket costs.

- For 75 first-time attacks, total shareholder loss is $104 billion.
- Direct out-of-pocket costs (investigation and remediation, penalties, etc.) is only $1.2 billion.
- Would suggest that breaches are informative, either about the underlying cyber-risk or the firm's capacity to provide cyber security.

Baseline Model

Model: Timing, $t = 1$

1. Nature draws private type of firm $\{C, N\}$, with prior $p(C) = \mu_1$.

Model: Timing, $t = 1$

1. Nature draws private type of firm $\{C, N\}$, with prior $p(C) = \mu_1$.
2. Consumers choose whether to be active and level of data to share with the firm in $t = 1$.

Model: Timing, $t = 1$

1. Nature draws private type of firm $\{C, N\}$, with prior $p(C) = \mu_1$.
2. Consumers choose whether to be active and level of data to share with the firm in $t = 1$.
   - Active users choose data d_1 to maximize exp. utility $u(d_1, p_1)$.
   - No access fee charged.

Model: Timing, $t = 1$

1. Nature draws private type of firm $\{C, N\}$, with prior $p(C) = \mu_1$.
2. Consumers choose whether to be active and level of data to share with the firm in $t = 1$.
   - Active users choose data d_1 to maximize exp. utility $u(d_1, p_1)$.
   - No access fee charged.
3. $t = 1$: Normal-type chooses unobserved $e_1 \in [0, 1]$, at cost $C(e)$.
   - type C is non-strategic : $e^C = 1$ in both periods.

Model: Timing, $t = 1$

1. Nature draws private type of firm $\{C, N\}$, with prior $p(C) = \mu_1$.
2. Consumers choose whether to be active and level of data to share with the firm in $t = 1$.
   - Active users choose data d_1 to maximize exp. utility $u(d_1, p_1)$.
   - No access fee charged.
3. $t = 1$: Normal-type chooses unobserved $e_1 \in [0, 1]$, at cost $C(e)$.
   - type C is non-strategic : $e^C = 1$ in both periods.
4. End of $t = 1$: a data breach may occur.
   - $P(b|e_1) = \zeta + (1 - \zeta)(1 - e_1)$, where $\zeta > 0$.
   - Integrate over types to get $p_1 = p(\mu_1, e_1)$
5. All consumers observe whether a breach occurs or not.

Model: Timing, t = 2

1. Beliefs are updated to $p_2 \in \{p_n, p_b\}$ using Bayes' rule.

Model: Timing, t = 2

1. Beliefs are updated to $p_2 \in \{p_n, p_b\}$ using Bayes' rule.
2. $t = 2$: Consumers choose participation and data sharing again.
   - Based their posterior belief about prob. of breach in $t = 2$.
3. End of $t = 2$: data-breach occurs or not, and the game ends.

Posterior beliefs

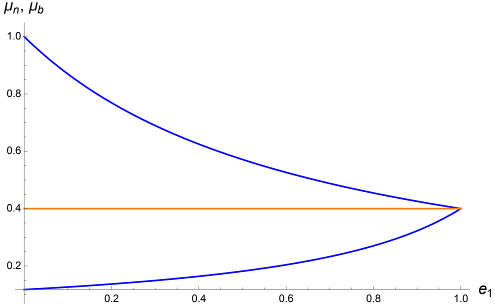

Figure 1: Posteriors μ_n (decreasing) and μ_b (increasing) as functions of first-period investment. As $e_1 \to 1$, outcomes become uninformative and they converge to the prior.

- Bayes' rule implies μ_n decreases in e_1 and μ_b increases.
- If $\zeta = 0$, perfect bad news: $\mu_b = 0$.

Model: Demand and revenue

Assumption 1: $u(d, p)$ quasi-concave in d, with $u_p \leq 0$ and $u_{d,p} \leq 0$.

- In each period, active consumers choose:

$$d^*(p) = \arg\max_d u(d, p) \tag{1}$$

- Assn 1 implies decreasing $d^*(p)$.

Model: Demand and revenue

Assumption 1: $u(d, p)$ quasi-concave in d, with $u_p \leq 0$ and $u_{d,p} \leq 0$.

- In each period, active consumers choose:

$$d^*(p) = \arg\max_d u(d, p) \tag{1}$$

- Assn 1 implies decreasing $d^*(p)$.

Assumption 2: Consumers have heterog. outside options, $\theta \sim F[0, 1]$.

- Mass of active users decreases in p.

Model: Demand and revenue

Assumption 1: $u(d, p)$ quasi-concave in d, with $u_p \leq 0$ and $u_{d,p} \leq 0$.

- In each period, active consumers choose:

$$d^*(p) = \arg\max_d u(d, p) \quad (1)$$

- Assn 1 implies decreasing $d^*(p)$.

Assumption 2: Consumers have heterog. outside options, $\theta \sim F[0, 1]$.

- Mass of active users decreases in p.

Assumption 3: $\Pi(p) := r(d^*) \cdot F(u(d^*, p))$, with $r'(d) > 0$.

- Revenue per consumer increases in d.
- $\Pi'(p) < 0$.

Model: Investment decision in $t = 1$

Taking consumers' investment beliefs, $\tilde{\mathbf{e}}_\mathbf{1}$ as *given*, the Normal type chooses e_1 to maximize:

$$\mathrm{T}\Pi = \Pi(p_1) - C(e_1) + P(b|e_1)\Pi(p_n) + (1 - e_1)\Pi(p_b)$$

Model: Investment decision in $t = 1$

Taking consumers' investment beliefs, $\tilde{\mathbf{e}}_{\mathbf{1}}$ as *given*, the Normal type chooses e_1 to maximize:

$$\mathrm{T}\Pi = \Pi(p_1) - C(e_1) + P(b|e_1)\Pi(p_n) + (1 - e_1)\Pi(p_b)$$

- Investment is purely retention driven: $e_2 = 0$ in any subgame of period 2.
- The firm's best-response to consumer beliefs $\tilde{e}_1$ is found by the foc:

$$(1 - \zeta)\big(\Pi(p_n) - \Pi(p_b)\big) = C'(e_1^{BR})$$

Model: Investment decision in $t = 1$

Taking consumers' investment beliefs, $\tilde{\mathbf{e_1}}$ as *given*, the Normal type chooses e_1 to maximize:

$$T\Pi = \Pi(p_1) - C(e_1) + P(b|e_1)\Pi(p_n) + (1 - e_1)\Pi(p_b)$$

- Investment is purely retention driven: $e_2 = 0$ in any subgame of period 2.
- The firm's best-response to consumer beliefs $\tilde{e}_1$ is found by the foc:

$$(1 - \zeta)\big(\Pi(p_n) - \Pi(p_b)\big) = C'(e_1^{BR})$$

- At equilibrium, beliefs must be correct, $e_1^{BR}(\tilde{e}_1) = \tilde{e}_1 = e_1^*$.

Proposition 1

There exists a unique Perfect Bayesian Eqm, $(e_1^*, \mathbf{p}^*, \mathbf{d}^*)$, of this game. It is separating, i.e. $e_1^* < 1$.

Welfare analysis of data-sharing

Welfare analysis of data-collection

In this section:

- A CS-maximizing planner can **ex-ante** mandate specific levels of d_2.
- Can condition d_2 on first-period outcomes, i.e. $d_2 \in \{d_n, d_b\}$.

Welfare analysis of data-collection

In this section:

- A CS-maximizing planner can **ex-ante** mandate specific levels of d_2.
- Can condition d_2 on first-period outcomes, i.e. $d_2 \in \{d_n, d_b\}$.

Starting from the unique "regulation-free equilibrium" $(e^*, \mathbf{p}^*, \mathbf{d}^*)$:

$$\frac{dCS}{d(d_b)} = \left[\frac{\partial CS_1}{\partial e_1} + \frac{\partial CS_2}{\partial e_1}\right]\frac{\partial e_1}{\partial d_b} + \underbrace{\frac{\partial CS_2}{\partial d_b}}_{=0} \tag{2}$$

Welfare analysis of data-collection

In this section:

- A CS-maximizing planner can **ex-ante** mandate specific levels of d_2.
- Can condition d_2 on first-period outcomes, i.e. $d_2 \in \{d_n, d_b\}$.

Starting from the unique "regulation-free equilibrium" $(e^*, \mathbf{p}^*, \mathbf{d}^*)$:

$$\frac{dCS}{d(d_b)} = \left[\frac{\partial CS_1}{\partial e_1} + \frac{\partial CS_2}{\partial e_1}\right]\frac{\partial e_1}{\partial d_b} + \underbrace{\frac{\partial CS_2}{\partial d_b}}_{=0} \tag{2}$$

Changes in either d_n or d_b affect CS via:

1. Direct effect on utility (not first-order)
2. Indirect effect on CS_1 via eqm security.
3. Indirect effect on CS_2 via distribution of posterior beliefs.

Will examine each term of the total derivative in sequence.

Effect on investment

Reminder: d_b = data to be shared in t=2 following a breach in t=1.

Lemma 1

At the unique equilibrium, a marginal increase in d_b decreases investment, $\partial e_1/\partial d_b < 0$.

1. d_b affects marginal profit of e_1 only via its impact on $t = 2$ profit following a breach.
2. When higher d_b increases profit following a breach, security incentives decrease.
3. At d_b^*, that profit is always *increasing* in d_b: Consumer-optimal sharing is **below** the ex post profit-maximizing value.

Welfare analysis: Signal jamming

What is the effect of e_1 on CS_2?

1. Changes frequency of breaches conditional on Normal type.
2. Changes posterior beliefs, and thus optimal consumer choices $\rightarrow$ not first-order.

Welfare analysis: Signal jamming

What is the effect of e_1 on CS_2?

1. Changes frequency of breaches conditional on Normal type.
2. Changes posterior beliefs, and thus optimal consumer choices $\rightarrow$ not first-order.

Lemma 2

The marginal impact of investment on CS_2, $\partial CS_2/\partial e_1$, is negative and increasing. If $\zeta > 0$, as $e \rightarrow 1$, it converges to zero.

Welfare analysis: Signal jamming

What is the effect of e_1 on CS_2?

1. Changes frequency of breaches conditional on Normal type.
2. Changes posterior beliefs, and thus optimal consumer choices $\rightarrow$ not first-order.

Lemma 2

The marginal impact of investment on CS_2, $\partial CS_2/\partial e_1$, is negative and increasing. If $\zeta > 0$, as $e \rightarrow 1$, it converges to zero.

Signal-Jamming Intuition: When facing a Normal firm, higher e_1 simply reduces the probability that consumers become aware, thus they choose sub-optimally high participation/data sharing in t=2.

Higher e_1 impedes learning about the firm's type.

Illustrating the Lemma

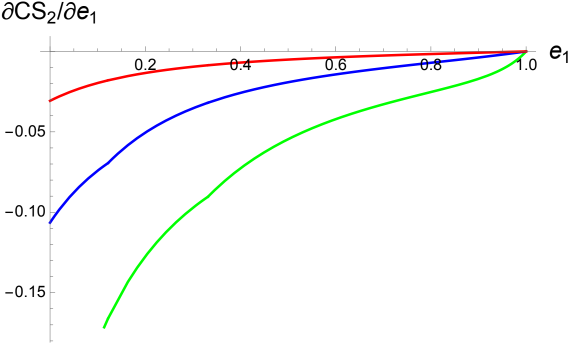

Figure 2: Illustration of Lemma 2: Greater investment impedes learning and decreases CS_2, but does so at a decreasing magnitude. In the Figure, as e varies, consumers adjust their beliefs and optimal decisions.

Red curve = high ζ; intuitively, lower impact of signal jamming when firm type is less informative.

Welfare analysis of data-collection

Putting everything together: Impact on CS_2 around regulation-free equilibrium:

$$\frac{dCS_2}{d(d_b)} = \underbrace{\frac{\partial CS_2}{\partial e_1}}_{(-)} \underbrace{\frac{\partial e_1}{\partial d_b}}_{(-)} + \underbrace{\frac{\partial CS_2}{\partial d_b}}_{=0} > 0 \tag{3}$$

Welfare analysis of data-collection

Putting everything together: Impact on CS_2 around regulation-free equilibrium:

$$\frac{dCS_2}{d(d_b)} = \underbrace{\frac{\partial CS_2}{\partial e_1}}_{(-)} \underbrace{\frac{\partial e_1}{\partial d_b}}_{(-)} + \underbrace{\frac{\partial CS_2}{\partial d_b}}_{=0} > 0 \tag{3}$$

Lemma 3

Starting at the initial equilibrium (e_1^*, p^*, d^*), the planner can increase CS_2 by **ex-ante** *imposing small caps on data-sharing for high-reputation firms, but not for low-reputation ones.*

From a CS_2 perspective, consumers share too little data with low-reputation firms, but give out too much data to high-reputation firms.

Welfare analysis of data-collection

Putting everything together: Impact on CS_2 around regulation-free equilibrium:

$$\frac{dCS_2}{d(d_b)} = \underbrace{\frac{\partial CS_2}{\partial e_1}}_{(-)} \underbrace{\frac{\partial e_1}{\partial d_b}}_{(-)} + \underbrace{\frac{\partial CS_2}{\partial d_b}}_{=0} > 0 \tag{3}$$

Lemma 3

Starting at the initial equilibrium (e_1^*,p^*,d^*), the planner can increase CS_2 by **ex-ante** *imposing small caps on data-sharing for high-reputation firms, but not for low-reputation ones.*

From a CS_2 perspective, consumers share too little data with low-reputation firms, but give out too much data to high-reputation firms.

- By imposing (ex-ante) data caps on data-sharing with high-reputation firms, the planner can achieve lower eqm e_1 and thus more learning.
- Will come at a cost of more frequent first-period breaches.

Welfare analysis of data-collection

Lemma 4

First period consumer surplus is a convex function of investment e_1. Total consumer surplus is also convex.

- High e_1^*: Higher participation and $d_1^* \to$ greater harm if a breach does occur.
- High e_1^*: Lower magnitude of negative signal jamming effect.

As a result, it is more likely that starting from equilibria with low e_1, increases in e_1 can potentially *decrease* total consumer surplus.

Fact: Across all (e, d_n, d_b) combinations, total CS is maximized when $e = 1$ and data-sharing is given by the ex-post optimal choices of consumers.

Total consumer surplus

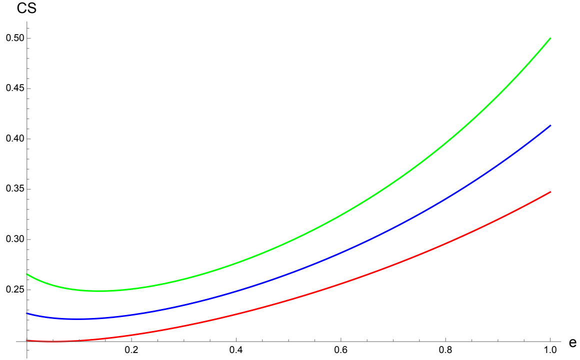

Figure 3: Total consumer surplus is a convex a function of e_1. Green curve corresponds to lowest value of ζ.

Literature review

- Economics of privacy **surveys**: Acquisti et al (2016), Goldfarb and Tucker (2023)
- **Strategic attackers**: De Corniere and Taylor (2022), Anhert et al (2023, also has moral hazard component), Fainmesser et al (2023)
- **Data storage and security choices**: Fainmesser et al (2023), Scheifert and Lam (2023)
- **Consumer learning**: Julien et al (2020), Toh (2018)
- Impact of cyber-attacks on firms (**empirical**): Kamiya et al 2021, Jamilov et al 2021, and many more.
- Other relevant theory work: De Corniere and Taylor (2021), Lefouili et al (2023), Markovich and Yehezkel (2023).
- **Impact of GDPR** on firm performance and outcomes (empirical): Aridor et al 2022, Johnson et al 2022.

Conclusion

Model:

- Reputation concerns incentivize firms to invest in cyber security.
- More data sharing raises revenue-per-consumer but also makes breaches more harmful.

Investment affects security, as well as *learning*.

- When consumers control ex-post data sharing, total CS might increase following changes that induce lower investment.
- When firms control ex-post data sharing, consumers benefit from imposing caps for both high and low reputation firms (didn't show today).

Thank you!

Additional Slides

Does this insight extend to $T = \infty$?

Take an example model with $T = \infty$:

- Firm lives forever, has private knowledge of its time-invariant type.
- Consumers have memories of 1 period. Once they become alive in period t, they immediately learn the security-outcome of period $t - 1$.
- Thus, when making their participation + data choices, their beliefs are either μ_n or μ_b.
- The firm chooses e in every period, and it is clear that there is an equilibrium in which it chooses the same e in each period.
- A fine that changes equilibrium e will affect both beliefs and security outcomes of each generation of consumers.

1. All previous results apply in this setting too! (equilibrium uniqueness requires suff. convex cost, even for $\zeta > 0$.)
2. Are there regions in which steady-state CS is **decreasing** in e?
   - Yes, if loss from reduced learning dominates security gains at $e = 0$.