Privacy Risk is a Function of Information Type Learnings for the Surveillance Capitalism Age

Ranjan Pal, *Member, IEEE,* Junhui Li, *Student Member, IEEE,* Jon Crowcroft, *Fellow, IEEE,* Yong Li, *Senior Member, IEEE,* Mingyan Liu, *Fellow, IEEE,* and Nishanth Sastry, *Senior Member, IEEE*

***Abstract*—In-app advertising is a multi-billion dollar industry that is an essential part of the current digital ecosystem, and is amenable to sensitive consumer information often being sold downstream without the knowledge of consumers, and in many cases to their annoyance. While this practice, in cases, may result in long-term benefits for the consumers, it can result in serious information privacy (IP) breaches of very significant impact (e.g., breach of genetic data) in the short term. The question we raise through this paper is: *Does the type of information being traded downstream play a role in the degree of IP risks generated?*. We investigate two general (one-many) information trading market structures between a single data aggregating seller (e.g., enterprise app) and multiple competing buyers (e.g., ad-networks, retailers), distinguished by mutually exclusive and privacy sanitized aggregated consumer data (information) types: (i) data entailing strategically complementary actions among buyers and (ii) data entailing strategically substituting actions among buyers. Our primary question of interest here is: *trading which type of data might pose less information privacy risks for society?* To this end, we show that at market equilibrium IP trading markets exhibiting strategic substitutes between buying firms pose lesser risks for IP in society, primarily because the 'substitutes' setting, in contrast to the 'complements' setting, economically incentivizes appropriate consumer data distortion by the seller in addition to restricting the proportion of buyers to which it sells. Moreover, we also show that irrespective of the data type traded by the seller, the likelihood of improved IP in society is higher if there is purposeful or free-riding based transfer/leakage of data between buying firms - simply because the seller finds itself economically incentivized to restrict the release of sanitized consumer data both, with respect to the span of its buyer space as well as in improved data quality.**

***Index Terms*—information privacy, strategic substitute, strategic complement, information market, Bayes Nash equilibria**

I. Introduction

Mobile applications (apps) are driving a major portion of the modern digital society, including business small and large as well as the state-of-the-art IoT/CPS systems. In-app advertising is an essential part of this digital ecosystem of mostly free mobile applications, where the ecosystem entities comprise the consumers, consumer apps, ad-networks, advertisers, and retailers (see Figure 1 (from [1]) for a simplified representation). As a social objective, a 'win-win' deal is desired between

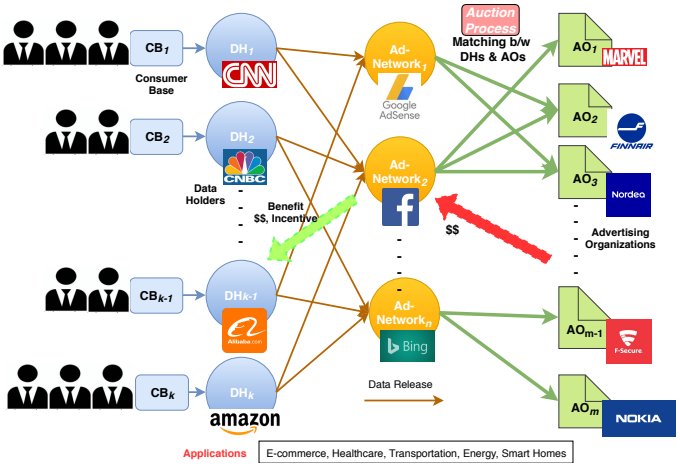

Fig. 1: Illustration of an Example Mobile In-App Ad Ecosystem

(a) the commercial interests of entities (e.g., enterprises, apps, databoxes) that aggregate and sell consumer data and those (e.g., ad-networks, retailers) that buy this data from the latter, (b) interests of consumer behavior targeting advertising firms, and (c) preserving consumer side information privacy (IP). The basic requirement for this 'win-win' ecosystem to exist in the first place, is the flow of personalized information from the consumer to the advertisers and retailers via the ad-networks (or directly from consumer to the advertisers/retailers) for effective/profitable ad placements, that subsequently motivate the latter to collect personal data about consumers via apps. As a popular example, the app version of *Evite.com* may sell lists of their consumers attending a party in a given location to targeted advertisers via ad-networks run by Google and Facebook. Similarly[1], the gene testing company *23andMe* might sell their clientele information directly to pharmaceutical companies in order for the latter to develop medical drugs. While this practice, in cases, may result in long-term benefits for the consumers, it can result in serious information privacy (IP) breaches of very significant impact (e.g., breach of genetic data leading to job/workplace discrimination) in the short term.

Research Motivation - In view of the just discussed contrasting effects of supply-side (consumer) sale downstream, one might be curious to know whether the type of information

R. Pal. M. Liu, and J. Li are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA. E-mail:*palr, mingyan, opheelia*@umich.edu

J. Crowcroft is with the Computer Laboratory, University of Cambridge, UK, and the Alan Turing Institute, UK, E-mail:jac22@cam.ac.uk

Y. Li is with the Department of Electronic Engineering, Tsinghua University, China. E-mail:liyong07@tsinghua.edu.cn

N. Sastry is with the Department of Informatics, King's College London, UK. E-mail:nishanth.sastry@kcl.ac.uk

[1]In general, the buyer set comprises (apart from Google and Facebook) an expansive, alphabet group of companies, from lesser-known organizations that help landlords research potential tenants or deliver marketing leads to insurance companies, to the quiet giants of data (**people search** firms such as Spokeo, ZoomInfo White Pages, etc.; **credit reporting** firms such as Equifax, Experian, etc.; **advertising and marketing** firms such as Oracle, Innovis, etc.

being traded downstream from the supply to the demand side in Figure 1 has an influence on the degree of IP risks generated in the ecosystem. This question naturally arises due to statistical correlations/dependencies that often exists between (i) the private and public attributes of an individual record of a database, and (ii) the individual records of a database. In both these cases, the database refers to a collection of multi-attribute records on app clients. Publicly known relationships between record holders can enable us to infer information about each of them, irrespective of their revelation preferences. Consequently, we aim to set out, first, adopting a mutually exclusive classification of supply side data types, and then following up with an analysis of the degree to which each type contributes to IP risks.

A. Research Contributions

We make the following contributions in this paper:

- We propose a consumer information trading market model between a *single* seller (a data holder in Figure 1) such as a mobile app that has in its possession aggregated personal data of its clients (consumers), and *multiple* strategic buyers (e.g., ad-networks, retailers). Our model formally captures mutually exclusive economic types (i.e., either being an economic substitute or an economic complement) of consumer data being traded, the commercial interests of the market stakeholders, as well as the social interests (e.g., privacy concerns) of the clientele of the seller. Through our model, we wish to study the optimal management of tradeoffs between information privacy (IP) welfare achieved in society and the commercial gains leveraged by the data selling services, as a function of the economic type of data being traded (See Section II).
- We analyze information trading markets for two mutually exclusive types of economic data being traded between strategic buying firms: (i) those entailing strategically substituting actions among buyers, and (ii) those entailing strategically complementing actions among buyers (see Section III for practical details). For each of these types, we analyze market equilibria in cases when the data seller, i.e., supplier, sells the same as well as different quality[2] of (privacy-sanitized) data to the strategic buyers. As our main result, we show that information trading markets with strategic substitutes data *mostly* (apart from situations when the substitutes nature between data buying firms is weak) pose less risks for IP in society at a market equilibrium, compared to markets where strategic buyer actions are over complements data. The primary intuition being (a) a section of buyers in substitutes settings are not economically incentivized to buy consumer data from a seller when they could (robustly) estimate such data for free from a statistical correlation analysis of the market environment, i.e., buyer priors, competing buyers, etc. - expecting this behavior, the strategic seller minimizes the correlation (thereby posing less risks of IP breaches) at market equilibrium in its products so as to maximize the number of buyers from the offered set, and (b) the seller, in the strong substitutes setting, does not gain in profit to sell data above a certain threshold fraction of strategic buyers, and in addition is incentivized to add extra noise to the data, enhancing chances of robust IP by making it hard to see through correlations (See Section III).
- We analyze the state of IP in society for single seller, multiple buyer information trading markets where the buying firms incorporate *purposeful* information exchanges (either via intentional free-riding channels, through paid third party sources, or collusion between firms) between competing buying firms to (robustly) infer the quality of consumer data floating in the market before they take their action. Under public knowledge of this assumption, in the substitutes settings, our analysis results in the finding (similar to those in Section III) that at a market equilibrium a profit-minded data seller is incentivized economically to (a) not offer to sell data to all possible buyers - but maximize the number of buyers in the offered set to actually buy data post offer, and (b) sell less precise consumer data to the chosen buyers compared to the preciseness in the case when there is assumed to be no information exchange between buying firms, so as to make a downstream firm's task difficult to estimate consumer data from inter-firm correlations. Consequently, the risk of IP breaches will be reduced (see Section IV).

[2]A difference in base quality can arise when a seller can sell two different types of data packages, e.g., basic and advanced (selling data on more attributes), on the same data it has about its cleintele, at heterogeneous prices.

II. Information Trading Model

In this section, we describe our information trading model (based on [2])[3] between a single downstream seller of aggregate cleintele data and multiple downstream strategic buyers, aligned with the ecosystem illustrated in Figure 1. Throughout this paper we use the terms 'seller' and 'data holder' (DH) (See Figure 1) interchangeably. We organize this section in four parts. First, we provide a qualitative overview of the mobile app ecosystem. Second, we model the data seller. Third, we model the strategic buyers. Finally, we model the competitive trading game between the DH and the downstream buyers.

A. Qualitative Overview of the Ecosystem

We provide an exemplified overview of our proposed ad-ecosystem shown in Figure 1 using mobile e-commerce applications as a representative example. Consumer apps from e-commerce websites hold, i.e., collect data/information of their clients that include contact details (phone), demographics (ZIP code), *do not call* flags, type of product in their inventory, etc. This data, for each client, is stored as a record of attributes [3] in an aggregate client database hosted/owned by the app. An ad-network, usually run by a search engine, hosts a market-place where suppliers (the consumer apps) sell "privacy-sanitized" versions (e.g., anonymous versions) of their respective databases. Businesses that sell their products to consumers via e-commerce apps form the demand side of this

[3]We re-use some notation for the purpose of consistency.

market place and are interested to buy records of anonymous consumers from the latter, that hosts a plethora of databases from multiple consumer e-commerce apps associated with it. The bought data by the businesses is a result of filtered search on database attributes, and are processed and AI/ML-analyzed [4] by the advertising/marketing wing of such businesses to target appropriate consumers. Each ad-network is a market-place that matches business ads with apps who show the former to their customers/clients. The matching decisions are a function of (a) quality (the number of attributes per record along with the degree of privacy-preserving perturbation on these attributes) of consumer data requested by businesses, (b) quantity (number of such records) bought, and (c) price-bids per record forwarded by the businesses to the market place.

B. Modeling the Data Seller

We consider a data seller (e.g., mobile app) having access to aggregate consumer data from its client base. This data, parts of it that is assumed to be private, could be a database. As popular practical examples, the firm *BookYourData* (BYD) offers downstream buyers ready-made lists of contacts of business individuals across different industries, job titles, job functions, and job levels. A record in a list consists of contact information such as name, email address, job function, job department, country etc. The organization *SalesLead* (SL) maintains a variety of datasets of American businesses in the form of profession-based lists and state/province-based lists - the *Accountant Sales Leads* dataset contains records of US-based accountants, whereas the *Alabama Sales Leads* dataset contains records of different businesses (accountants, real-estate agents, etc.) based in Alabama. Each record in a dataset consists of contact information such as mailing address, geo-location, email address, phone number, etc. As another major example, the telemarketing company *TelephoneLists* specializes in offering its buyers phone lists as datasets that consists of information on consumers (contact details, demographics, etc.) as well as businesses (number of employees, sales, volume, etc.) in the US and Canada.

We assume that the seller has access to a private signal z of its client database θ with an accuracy level characterized by κ_z. The database under consideration could consist solely of private or public consumer attributes, or even a mix. We note that our analysis is also applicable to signals that do not have a structure of a database. Without loss of generality we consider a database signal in this paper. Mathematically, the relation between z, θ, and κ_z can be represented as:

$$z = \theta + \zeta,\ \zeta \sim \mathcal{N}(0, \frac{1}{\kappa_z}),$$

where $\mathcal{N}(0, \frac{1}{\kappa_z})$ is the imprecision level or the noise term associated with θ, and is assumed to be normally distributed[4]. Note that the noise is applied on all applicable attributes across all records. This imprecision can arise due to the unavailability, vagueness, or falsity of certain attribute information in the database due to voluntary choice measures taken by a subset of clients to release correct information. Note here that $z = \theta$ is assumed to be the ground truth state (i.e., a state without any ambiguity, unavailability, or falsity) of the database and may or may not be accessible to the seller. To this end, we assume that θ is perceived to the seller as a random variable z with finite support and taking a continuous statistical distribution of the different possible ground truth states. Thus far, we have also assumed that the utility of the entire database is isomorphic to the utility of each entry in the database, i.e., assuming that statistical noise is the only quality-hit on entries of bought records, each entry has the same utility to the buyer as the entire database (under the assumption of no missing entries and that the buyer has access to either all his records or none of them) as all entries take noise from the same distribution. Our choice of using a normal distribution to model database imprecision arises out of (a) the need to apply the central limit theorem on the database attributes for a significant client population, (b) the necessity of analytical tractability, and (c) the popularity of the Gaussian distribution as a practical noise generation mechanism in information-theoretic modeling.

Given access to client database signal z, the seller has control over who to sell/offer z to and at what precision. More specifically, for a given $i \in I$, the seller decides to offer a signal s_i of its client database to buyer i that is mathematically represented as:

$$s_i = z + \xi_i,\ \xi \sim \mathcal{N}(0, \frac{1}{\kappa_\xi}),$$

where $\mathcal{N}(0, \frac{1}{\kappa_\xi})$ is the noise term associated with z, and is assumed to be normally distributed similar to the motivation behind modeling z. We assume that ξ_i (controlled by the seller/DH and heterogeneous among the buyers) is independent of z and denotes the degree of privacy sanitization to signal z obtained after applying privacy preserving technologies (e.g., differential privacy to preserve anonymity, information-theoretic privacy to preserve privacy of sensitive attributes - not necessarily anonymity) to it. Thus, more (less) the value of ξ_i, greater (lesser) the privacy (utility) of z to buyer i. The DH sells s_i to i at a price of p_i. The noise term added by the DH/seller may or may not be correlated among different buying firms. We model the pairwise correlation between database signal of buying firms i and j as the *maximal correlation*[5], $\rho_\xi \in [0, 1]$, between the noise elements of signals for i and j. The motivation to use this correlation measure is to account for non-linear dependencies between the noise elements that cannot be captured through popular measures such as the popular Pearson or Kendall correlation measures. Signal noise elements that are non-correlated are statistically independent. Thus, in terms of the ground truth θ of client information, s_i can be expressed as:

$$s_i = \theta + \eta_i,\ \eta_i \sim \mathcal{N}(0, \frac{1}{\kappa_s}),\ C(\eta_i, \eta_j) = \rho,$$

where $\kappa_s = (\frac{1}{\kappa_z} + \frac{1}{\kappa_\xi})^{-1}$ is the precision level of s_i and $C(\eta_i, \eta_j) = \frac{\kappa_\xi + \rho_\xi \kappa_z}{\kappa_\xi + \kappa_z}$ is the intraclass correlation[6][5] (ICC)

[4]For the purpose of analysis and tractability, we assume that all attributes in the dataset can be noisified using continuous statistical distributions.

[5]The correlation here is between two classes of noise, one coming from ζ and the other from the ξ's.)

TABLE I: Table of Important Notations

θ	true realization of client database
ζ	noise level associated with θ
z	data seller's realization of θ
κ_z	accuracy level of signal θ
s_i	client signal offered for sale to buyer i
ξ_i	noise level associated with s_i
κ_ξ	precision level of signal z
ρ_ξ	pairwise MC matrix between signals of buyers
ρ	ICC between noise elements of signals of buyers i, j
κ_s	precision level of s_i
$v(a_i)$	utility of taking action a_i by buyer i
λ	fraction of buyer population offered a selling contract
x_i	prior estimate of s_i by buyer i
κ_x	precision level of x_i
π_i	payoff/profit function for data seller
Π	expected profit made by the data seller
β	degree of strategic complementarity
$\alpha_0, \alpha_1, \alpha_2, \gamma$	model constants

between noise elements of database signals for buying firms i and j.We also have that the signals sold by the seller cannot be more precise than what she gathers from its clients, i.e., $\kappa_s \leq \kappa_z$. This evidently follows because the seller will usually add some noise on top of z before selling s. In the best case $z = s$, i.e., no noise added to z, and we will have $\kappa_s = \kappa_z$. Finally, we assume that the seller pre-decides at market equilibrium which buying firms to offer consumer data to after which data trading is done between the seller and the buyers to potentially arrive at a market equilibrium. Note that, due to *ex-ante* symmetry, this scenario is isomorphic to the case when the seller pre-announces at market equilibrium its price and quality of data it is going to offer to individual buyers, after which trading is done to arrive at an equilibrium setting of buyers who decide to buy and those opting out.

C. Modeling the Buying Organizations

We consider a continuum of buying firms indexed by $i \in [0, 1]$ for two particular reasons: (i) there could be many[6] buying organizations having market power ranging from small to big, and (ii) working with a continuous mass of firms given a potentially large number of them appropriately calls for a non-discrete analysis w.r.t. the number of firms, though it makes the analysis relatively more challenging and complete when compared to a discrete setting.

On a broader intuitive level, a buying firm's profit function should ideally comprise three components: the first component should indicate the individual benefit to buyer i on buying data related to θ with a particular degree of precision. The second component should indicate the individual benefit to buyer i conditioned on the aggregate action of all buyers - a higher benefit (due to positive externality) being accrued if competitors invest in high quality/precision data (as better inferences can be made from competitor data with lesser investments). The third and last component should indicate the cost incurred by buyer i to adopt strategy a_i - usually a quadratic function in many economic settings.

[6]According to a new Vermont law, data brokers are required to register with the Secretary of state - according to them there are roughly 121 small brokers just in the USA as a conservative estimate. *Source* - FastCompany.com

The cost function, i.e., the third component, according to traditional modeling practices in micro-economic theory [7], should ideally obey regularity conditions the cost function must satisfy the following regularity conditions: *continuity, symmetry, linear homogeneity in prices, monotonicity in prices and outputs*, and *concavity in prices*. In addition, the cost function should be of a flexible functional form that can be used to approximate any twice-differentiable cost function that may result empirically. The trans-logarithmic and quadratic cost functions are two function types that satisfy these conditions. Flexibility notwithstanding, the trans-logarithmic functional form has a major limitation - its inability to deal with zero levels of outputs that is a theoretical/pathological possibility in our model setting. The quadratic cost model specification offers a better alternative in this regard - exhibiting the flexibility of the translog function while conforming to the properties of economic theory.

More formally, each buying firm i takes an action $a_i \in \mathbf{R}$ to maximize its profit that is mathematically expressed as:

$$\pi(a_i, A, \theta) = \alpha_0 v(a_i) u(\theta) + \alpha_1 v(a_i) v(A) - \frac{\alpha_2}{2} v(a_i)^2, \quad (1)$$

where $A = \int_0^1 a_i di$ denotes the aggregate of individual actions a_i taken by the buying firms, and $\{\alpha_0, \alpha_1, \alpha_2\}$ are exogenously given constants. Here, action a_i in practice might reflect (a) the degree of precision of the data bought, with a higher precision implying lesser noise perturbation and vice-versa, or (b) price at which per unit of information is bought, with higher prices indicating less noisy data and vice-versa. Both (a) and (b) generalize the popular Bertrand and Cournot market competition structures prevalent in mainstream microeconomics to trade goods (in our case digital information). Consequently, the action of a buyer is influenced by the weighted sum of its estimated prior of the ground truth (denoted by θ), and current signal from the seller that results in an updated degree of signal precision for the buyer. This update is made post the move by the seller, and dictates the actual action for the buyer. Action variables can also be non-scalar (e.g., privacy-related supply functions [1][8][9]) that indicate mathematical functions that specify monotonic preferences on a_i, instead of a single value - with higher preferences less noisy data. Such action spaces are important to reach approximately optimal trading market outcomes when it is not possible to arrive at optimal ones. However, for simplicity purposes, we do not consider such action spaces in this work. $u(\theta) \in \mathbf{R}$ is a random variable representing the utility function to the buyer of a probabilistic-ally uncertain (only to a buyer) ground truth state. i.e., the original database at the seller's disposal, of the consumer database and taking the same statistical distribution as θ. In practice, a buyer will not have access to the ground truth (due to privacy perturbations by the seller pre-sale), but will have a probabilistic belief about the same, and consequently its utility will be driven by this belief and hence a r.v. $v(\cdot) \in \mathbf{R}$ is a utility function to the buyer of the strategic action it takes, and $v(A) = \int_0^1 v(a_i) di$. Here, we assume for analytical simplicity and tractability purposes that (a) u's and v's are the same for all firms, and (b) $v(A)$ is linearly separable in a_i as just shown through the

expression for $v(A)$. In addition, the linearly separable form of $v(A)$ is one of the feasible ways to capture the practical notion behind $v(A)$. Referring back to strategy variable a_i per buyer (after it decides to opt in or opt out of trade with the data seller), we model it in our paper to be the precision of bought data/signal that is expressed as an appropriately weighted linear combination of signal precision regarding θ (see Section III for details).

We assume, partly for simplicity, that all buying firms hold a proper uniform distribution[7] to be the common prior on θ to reflect the situation that buyers have no bias on particular states of θ. However, With respect to s_i sold by the seller to firm i, we assume the latter has access to a (non-uniform) prior version of s_i in the form of x_i that is mathematically expressed via the following $x_i = \theta + \epsilon_i$, $\epsilon_i \sim \mathcal{N}(0, \frac{1}{\kappa_x})$, where κ_x denotes the precision of the prior version of the data signal s_i sold by the seller, as observed by i prior to trading. In practice, such priors can be arrived at through market study of related published consumer information from similar businesses. We assume that ϵ_i's are independent across the firms.

Given the profit function of firm i in Equation 1, we let

$$\beta = -\frac{\frac{\partial^2 \pi}{\partial a \partial A}}{\frac{\partial^2 \pi}{\partial a^2}} = \frac{\alpha_1}{\alpha_2}, \tag{2}$$

be the degree of strategic complementarity in firm i's actions. The case when $\beta > 0$ corresponds to the scenario when the buying firms' actions are *strategic complements*, i.e., the higher the actions (i.e., precision of bought signal) of other firms are, higher is the benefit to firm i to buy a signal of higher precision [reflected in the second component of Equation 1]. In practice this complementarity may arise when the seller's information (e.g., some personal consumer attributes important for profitable targeted advertising and which cannot be obtained/ robustly estimated without pay from the market environment) is important enough for the interests of a buying firm's profit motives and improvement in market share/power, and thus it is incentivized to invest in such information when other competitors do. *A popular example of such information is* Fitbit *data which is in high demand by market competitors keen to take advantage alongside their strategic partners.*

In a similar fashion the case when $\beta < 0$ corresponds to the scenario when the buying firms' actions are *strategic substitutes*, i.e., the benefit to firm i of taking a higher action (i.e., precision of bought signal) decreases with the aggregate action A [also reflected in the second component of Equation 1]. In practice this *substitutes* characteristic may arise when the seller's information (e.g., either some personal consumer attributes or public attributes, estimates of both which could possibly be gathered from the market environment (cheaply through second or third party sources) without buying from the seller) is not necessary enough for the buyer side to pay and contribute to its interests of increasing market share/power. Consequently, due to high positive externalities from competitor acquired data that arise through statistical correlations, a buyer is not incentivized to invest in such information when other competitors might. *An example of such information is data gathered for free by Google/Facebook when we login to a website using the former's social account.* The case when $\beta = 0$ corresponds to the scenario in which buying firms face no strategic interactions (the domain of non-replacable data of no/less commercial use). We also assume that $\alpha_2 > \max\{2\alpha_1, 0\}$ which makes β lie in the interval $(-\infty, \frac{1}{2})$ in order to guarantee that buying firm i's profits are strictly concave in a_i so as to reflect the property of decreasing marginal returns.

[7]Suppose that θ i distributed according to a Gaussian distribution with mean 0 and variance σ_θ^2. By letting $\sigma_\theta \to \infty$, we obtain a statistical distribution with full support over $(-\infty, \infty)$ that, in the limit, assign the same probability to all intervals with the same Lebesgue measure.

D. Modeling The Information Trading Game

Once the seller and individual strategic buyers have their own prior estimates z and x_i's of θ respectively, the former opts to sell s_i's to the buyers. Consequently, the goal of the seller is to advertise a 'take-it-or-leave-it' offer $(\kappa_\xi, \rho_\xi, p_i)$ to the fraction λ of buying firms. The buying firms $i \in [0, \lambda]$ observe the advertised contract and have the freedom to decide whether to accept ($b_i = 1$) the contract or to reject ($b_i = 0$) it. After the decision taken by each of the buyers, there is the trading competition subgame in which each strategic buyer chooses their action a_i to maximize their profit function. Taking into account the rational mindset of the selfish buyers to choose the optimal action on being offered a contract, that maximizes their payoffs, the rational seller designs the optimal contract offer. Thus, we have a dynamic game setting, like in [2]. Note that while buying firm i opting in needs to map (x_i, s_i) to a_i, a buyer i opting out needs to map x_i to a_i.

In theory, the information trading ecosystem desires a stable/equilibrium operating state where (i) the seller chooses parameters $(\lambda, \kappa_\xi, \rho_\xi, \{p_i\}_{i\in[0,\lambda]})$, (ii) each data buyer makes acceptance/rejection decisions $b_i \in \{0, 1\}$, (iii) a posterior belief μ_i on θ is evaluated by each strategic firm i, and (iv) action a_i is taken by each buyer i such that the following conditions hold:

1) the seller chooses $(\lambda, \kappa_\xi, \rho_\xi, \{p_i\}_{i\in[0,\lambda]})$ to maximize its own profit;
2) each firm i accepts the seller's offer if doing so maximizes its profit;
3) for each firm i, the μ_i posterior estimate of θ is obtained via the Bayes rule conditioned on the information set;
4) given its posterior belief, each firm i maximizes expected payoffs in the competition subgame, taking the strategies of all other firms as given;
5) the aggregate action A is consistent with individual firm-level actions.

The operating state that satisfies the above five conditions is popularly known to be a *perfect Bayesian equilibrium* state in game theory, derived as the outcome of a dynamic Bayesian game of incomplete information [7].

III. Market Analyses

In this section, we analyze different variations of information trading markets in view of our proposed market trading model of Section II. More specifically, we (i) state the practical relevance of market variations and for each, (ii)

mathematically characterize the seller's optimal strategy to individual strategic buyers, and (iii) lay down the commercial and societal implications of information trading at market equilibria. To this end, we structure the section in three parts. In the first and second part we analyse the competition game between buying firms once the seller has decided the fraction of buyers to whom to offer a contract to, and then derive market equilibrium of the resulting competition - depending on the economic type of the data being traded (see Sections III-A and III-B, respectively). In the third part, we extend this analysis for seller products having quality gradations, in contrast to the case of homogeneous quality trading analysed in Section III-A.

Prior to delving into the details of the three parts aforementioned, we would like to emphasize that the games entailed by the oligopoly are not finite games, i.e., the strategy space consists of real numbers. Thus, the existence of a mixed-strategy market Nash equilibrium cannot be guaranteed via the seminal Nash's theorem [10] - however, both a pure strategy market Nash and a mixed-strategy Nash can be shown to exist via a theorem proposed by Glicksberg [11], and a theorem proposed by Fan, Debreu, and Glicksberg [12], respectively, both of which does not restrict the condition of equilibrium existence to finite strategy spaces. In our work, we will evaluate the more practically viable pure strategy market Nash equilibrium of our oligopoly game structures.

A. Buying Oligopoly Game - Buying Firm Strategy

Consider the situation where the data seller has already decided to offer contracts to a fraction λ (the strategy behind selection of the optimal λ is discussed in Section III.B) of the buyers, and that a fraction $l \leq \lambda$ of buyers eventually opt in. The primary question of interest here is: *what is the optimal strategy for the strategic buying firms*?. Before we answer this question, it makes sense to explain the meaning of strategy for a buying firm.

The Strategy for a Buying Firm - Each buying firm i has access to a private prior x_i of θ, and is sold s_i (if it opts in) by the seller. x_i is generally obtained from second or third party sources, either for free or at a price much lesser than what it needs to pay for s_i. Each firm who is offered s_i either (a) operates with x_i (and its posteriors) without buying s_i, (b) buys s_i and disregards x_i showing no confidence in its private prior, or (c) effectively combines x_i and s_i showing a share of confidence in both its private prior and paid signal. The strategic decision firm i needs to make encompassing (a), (b), and (c) above is: *how much weight to put on x_i as compared to s_i*?. The weighted combination of x_i and s_i subsequently acts as the action a_i for each buyer i. The answer to the previous question depends on the precision levels (quantity) of x_i and s_i, in addition to l - the number of buying firms opting in the contracts offered by the buyer, and ρ the pairwise ICC correlation matrix. l contributes to the importance of the data under trade to buyers, whereas ρ reflects the similarity of bought data among the firms. We have the following result, based on [2] characterizing the equilibrium actions of the strategic buyers, the proof of which is in the Appendix [13].

Theorem 3.1: The buying oligopoly game has a unique Bayesian Nash equilibrium. The market equilibrium strategic actions of each buying firm (assuming they opt for a linear combination of their private prior and bought signals) is given by:

$$a_i = \begin{cases} \gamma[(1-\omega)x_i + \omega s_i], \text{ if} i \in [0, l] \\ \gamma x_i, \text{ if} i \in [l, 1], \end{cases}$$

where $\omega = \frac{\kappa_s}{(1-\beta l\rho)\kappa_x + \kappa_s}$, *and* $\gamma = \frac{\alpha_0}{\alpha_2 - \alpha_1}$.

Theorem Implications and Intuition - The theorem states that at market equilibrium, firm i puts a weight split of ω and $1-\omega$ to the private prior and bought signal respectively, where ω is a function of l and ρ. Importantly, the weight firm i assigns to s_i, irrespective of ρ and l, increases with β - the degree of strategic complementarity. The intuition behind this result is that buying firms have a significant incentive to implicitly but non-cooperatively coordinate with one another (via a market 'eye' on other firms) with increasing β. This is simply because higher β indicates consumer information is commercially useful/vital enough (e.g., cyber-hygiene parameters of insurance clients) but not comparably 'easy' to publicly estimate (via inputs from third party sources) in a robust fashion. Thereby, each firm wants to invest and put more weight on bought data when compared to their "not so informative" priors, and improve their business/advertising prospects in order to increase market power/share. In the situation when β = 0, the optimal strategy of buying firms is only dependent on the precision metrics κ_x and κ_s, of the prior and bought signal respectively, and independent of l and ρ. This is intuitive given that without any complementarity (or substitutability) effect, it does not matter how many competitors buy data from the seller, or the strength of ICC between sold signals. Now keeping β fixed, we observe that the weights assigned to bought signals is monotonic in l and ρ. More specifically when $\beta > 0$ (the complementarity case), the weights increase in l and ρ - since as discussed above, complementarity along with increasing l commercially incentivizes firms to invest in bought signals. The weights also increase in ρ, i.e., correlation among firms; this might seem counter-intuitive but is perfectly justified in the context of $\beta > 0$ scenarios where firms find it incentive compatible to invest in the more informative bought data than rely on correlated data 'floating' around. A similar *converse* argument holds when $\beta < 0$ (the substitutes case), and we omit it for purposes of brevity.

B. Buying Oligopoly Game - Selling Firm Strategy

In this section, we discuss the strategy of the selling firm at a market equilibrium point. More specifically, we investigate at the unique market equilibrium (established in Section III-A), the quality of consumer data sold, and the fraction to whom a selling offer is made. In addition, we discuss the practical social implications of our analysis with respect to privacy welfare in society. We consider the scenarios of data being strategic complements as well as being strategic substitutes. Moreover, the analysis in this section is built upon framework in [2], and based on the assumption that the sellers are unaffected about privacy concerns of its consumers. i.e., are not subject to considerable regulatory fines - something quite

common to many mobile app firm businesses around the world *today*. *We relax this assumption in Section IV to account for the case when data sellers do consider privacy breach impacts to their business prospects* - something consumer data selling firms might have to think of in the near future, especially in the wake of heavy fines to be exercised (or already in fashion) on firms in the wake of policies such as the GDPR.

1) The Case of Strategic Complements: In the case when $\beta > 0$, i.e., the selling data exhibit strategic complementarities among the buying firms, we have the following result, the proof of which is in the Appendix [13].

Theorem 3.2: At the unique Bayesian Nash equilibrium of the buying oligopoly game with $\beta > 0$, the optimal strategy for a data seller is to sell an undistorted version of consumer data, $\kappa_s^* = \kappa_z$, to its downstream buyers and offer this contract to the entire set of the firm population, i.e., $\lambda^* = 1$. In return, the seller makes an expected profit of

$$\Pi^* = \gamma^2(\frac{\alpha_2}{2})(\frac{\kappa_z}{\kappa_x})\frac{\kappa_z + \kappa_x}{[(1-\beta)\kappa_x + \kappa_z]^2}. \tag{3}$$

Theorem Intuition and Practical Implications - The result simply states that at trade/market equilibrium, the data seller should offer to sell z to all buying firms in the pool without adding further statistical perturbations for privacy-preserving interests. Clearly, from the seller side, the result is intuitive given the fact that there is no regulatory liability/punishment on the data seller to preserve privacy of consumer data, and hence it finds it optimal to sell an unperturbed version of the data to gain maximum profits. From the buyer side, since data exhibit strategic complementarities, their surplus increases after the trade with the seller. As a consequence of the complementarity property of data sold to buyers, the seller is further incentivized to increase their price to improve their profit margins, as buyers find the data important for their businesses. On the seller profit front, we observe at equilibrium that the profit margin increases with κ_z, the precision of clientele information available to the seller - the buyers find it more useful and hence want to invest in it, and the seller can charge appropriately. However, the profit margins also decrease with increase in the precision, κ_x, of the priors the buyer population has on θ - simply (and intuitively) because it prevents a significant fraction of the population from buying information from the seller.

Privacy Implications for Society - In the strategic complements trading context without sufficient regulatory penalties, consumer privacy might be hampered if the data sold under consideration has private attributes, as there is no incentive for the seller to perturb z. In the case only public attributes of consumers are under sale, the risk of breaching consumer privacy depends on the correlation between the public attributes sold and the private attributes. Greater the correlation, greater the risk.

2) The Case of Strategic Substitutes: In the case when $\beta < 0$, i.e., the selling data exhibit strategic substitutes among the buying firms, we have the following result, the proof of which is in the Appendix [13].

Theorem 3.3: At the unique Bayesian Nash equilibrium of the buying oligopoly game with $\beta < 0$ and $-(1+\frac{\kappa_z}{\kappa_x}) \leq \beta < 0$, the optimal strategy for a data seller is to sell an undistorted version of consumer data, $\kappa_s^* = \kappa_z$, to its downstream buyers and offer this contract to the entire set of the firm population, i.e., $\lambda^* = 1$. In return, the seller makes an expected profit of

$$\Pi^* = \gamma^2(\frac{\alpha_2}{2})(\frac{\kappa_z}{\kappa_x})\frac{\kappa_z + \kappa_x}{[(1-\beta)\kappa_x + \kappa_z]^2}. \tag{4}$$

Theorem Intuition and Practical Implications - First, note that $\beta < 0$ (see below for a formal expression) in this setting, denotes the degree of substitutability of consumer information, and grows on the negative scale. The result simply states that under a *weak enough* substitutability of consumer data desired by downstream buying firms, at trade/market equilibrium, the data seller should offer to sell z to all buying firms in the pool without adding further statistical perturbations for privacy-preserving interests - *the same result as in Theorem 3.2 of the complements case*. As above-mentioned, the intensity of substitutability between firms is characterized by $\beta = \frac{\delta-1}{2\delta} < 0$ and is decreasing in $\delta \in [0, 1]$ (i.e., behaves increasing like complements). Consequently, in practice, higher the value of δ, more is the variance in the mindset of buyers to significantly rely on market environment estimate of θ, i.e., the individual priors. As a result, intuitively the seller (with increasing δ) finds it profit optimal to sell the best quality of consumer data at its disposal as buyer surplus is boosted in the weak substitutes case. On the seller profit front, we observe at equilibrium that the profit margin is lower in the weak substitutes case (due to a higher value of $1-\beta$), when compared to the complements case. This is due to the fact that in the weak substitutes case, buying firms have significant variance in the weights they allocate to the bought signal and their prior, unlike in the complements case where the variance is not as much. The variation of seller profit margin Π^* at trade equilibrium with κ_z and κ_x follows similar trends as in the complements case and follows a similar intuition.

Privacy Implications for Society - We just showed that the case for weak substitutes is similar to that of complements. The privacy implications are similar. i.e., without sufficient regulatory penalties, consumer privacy might be hampered if the data sold under consideration has private attributes, as there is no incentive for the seller to perturb z. In the case only public attributes under sale, the risk of breaching consumer privacy is positively monotonic with the correlation between the private and public attributes of consumer data.

For the substitutes case when $\beta < -(1+\frac{\kappa_z}{\kappa_x})$, we have the following result, the proof of which is in the Appendix [13].

Theorem 3.4: At the unique Bayesian Nash equilibrium of the oligopoly game with $\beta < -(1+\frac{\kappa_z}{\kappa_x})$, the optimal strategy for a consumer data seller is to sell a distorted version of consumer data, $\kappa_s^* < \kappa_z$, to its downstream buyers and/or offer this contract to a proper fraction of the firm population, i.e., $\lambda^* < 1$. More specifically, the value of κ_s^* and λ^* is given by the solution of the following equation:

$$(\kappa_z + \beta\lambda^*\kappa_s^*)\kappa_x + \kappa_z\kappa_s^* = 0. \tag{5}$$

In return, the seller makes an expected profit of

$$\Pi^* = -\gamma^2(\frac{\alpha_2}{2})(\frac{\kappa_z}{4\beta\kappa_x^2}). \tag{6}$$

Theorem Intuition and Practical Implications - The result simply states that under a *strong enough* substitutability of consumer data desired by downstream buying firms, at trade/market equilibrium, the data seller should offer to sell z to a subset of buying firms in the pool by adding further statistical perturbations for privacy-preserving interests. This is clear from Equation 5 which suggests that the fraction λ of buying firms that the seller offers to trade with and the information precision level, κ_s, are substitutes, i.e., as the seller prefers an increased λ, it is incentivized to increasingly distort the sold data. Intuitively, this follows because when data are sufficient substitutes an increased λ implies that the buying firms can implicitly coordinate with competing firms in the market, i.e., the market environment, and can robustly estimate consumer information without buying. This is precisely what the seller wants to discourage in order to increase sales, and thereby the increase in noise to promote sales. If the seller were to make $\kappa_s^* = \kappa_z$, the substitutes property of data will make a considerable number of firms robustly estimate consumer information without opting in a trade contract with the seller - something not aligned with the profit interests of the latter. On the seller profit front, we observe at equilibrium that the profit margin decreases with $|\beta|$ - the degree of the absolute value of strategic substitutability of the actions of buying firms. This is because a firm's incentive to buy data decreases with increased substitutability nature of data on the selling market. Finally, observe that the threshold $-(1+\frac{\kappa_z}{\kappa_x})$ at which the seller finds it optimal to limit her market share λ and/or strategically distort the consumer information, is decreasing in $\frac{\kappa_x}{\kappa_z}$, implying that the more informed the seller is relative to its downstream customers, the more likely she will able to exploit her informational advantage by selling it to the entire firms without distortion.

Privacy Implications for Society - The privacy implications for the strong substitutes case favor a comparatively better "privacy-friendly" society when compared to the complements case. This, because the sold data is distorted and devoid of moderate to high correlations and thus it is much hard work to breach consumer privacy. One should note here that our notion of privacy is mainly from the point of seeing through correlations without spending money. It is quite possible that downstream firms do invest to not rely on correlations, specially for important substitutes data, in which case good quality consumer data will reach them and consumer privacy might be under threat. However firms would now need to pay. A regulatory action is called for here to set trade prices to enforce the *data minimization principle* - only collecting data that is necessary, beyond which it should become cost prohibitive for buying firms.

C. The Case of Selling Differentiated Quality Downstream

In this section, we analyze the case when the data seller provides different packages to its downstream sellers related to its clientele information. This scenario is a practical use case since different buying firms might often be interested in package gradations at heterogeneous prices. As an example, differentiation may be in terms of the type and number of attributes sold downstream, in addition to the quality of information traded. We emphasize here that we are in the setting where the seller does not face a significant penalty (regulatory or otherwise) on adverse effects of selling consumer data to downstream buyers. Mathematically, we assume that the seller offers (κ_{s_i}, p_i) to each firm $i \in [0,1]$, specifying the precision level κ_{s_i} and price p_i. It is evident that $\kappa_{s_i} \leq \kappa_z$ for all i, i.e., the seller cannot provide the buyers with a precision level of their clientele data that is greater than its own private estimate z. We have the following result, based upon [2], characterizing the market equilibrium trade parameters, the proof of which is in the Appendix [13].

Theorem 3.5: At the non-unique Bayesian Nash equilibrium of the buying oligopoly game with $\beta \geq -(1+\frac{\kappa_z}{\kappa_x})$, the optimal strategy for a consumer data seller is to sell an undistorted version of consumer data, $\kappa_s^* = \kappa_z$, to its downstream buyers and offer this contract to the entire set of the firm population, i.e., $\lambda^* = 1$. In return, the seller makes an expected profit of

$$p^* = \gamma^2 (\frac{\alpha_2}{2})(\frac{\kappa_z}{\kappa_x}) \frac{\kappa_z + \kappa_x}{[(1-\beta)\kappa_x + \kappa_z]^2}. \tag{7}$$

If $\beta < -(1+\frac{\kappa_z}{\kappa_x})$, the optimal strategy for a consumer data seller is to sell clientele data at a precision $\kappa_{s_i}^*$ to buying firm i, where $\{\kappa_{s_i}^*\}_{i\in[0,1]}$ solves

$$\int_0^1 \frac{\kappa_{s_i}^*}{\kappa_x + \kappa_{s_i^*}} di = -\frac{\kappa_z}{\beta \kappa_x}, \tag{8}$$

at a price $p_i^* = \gamma^2 (\frac{\alpha_2}{2}) \frac{\kappa_{s_i}^*}{4\kappa_x(\kappa_x + \kappa_{s_i^*})}$.

Theorem Intuition and Practical Implications - The case of $\beta \geq -(1+\frac{\kappa_z}{\kappa_x})$ generalizes to the strategic complement setting and also the substitutes setting when the substitutes are weak (β negative but near to zero) - seller offers the highest precision consumer data available to the buyers. The intuition is the same as mentioned above for these scenarios - primarily that the 'complements' nature of consumer data incentivizes both the profit minded seller, and also the strategic buyers to trade with the highest quality data available. On a similar note, for the case when $\beta < -(1+\frac{\kappa_z}{\kappa_x})$, it is not incentive compatible for the seller to offer the highest precision consumer data at its disposal to downstream buyers, and that too to the entire population of buyers. It is easy to see that $\kappa_{s_i}^* = \kappa_z$ does not satisfy optimality condition (8), for all buyers i. The intuition behind this result is that providing high quality data increases the correlation in the firms' actions, which in turn reduces their profit when the actions are strong strategic substitutes. Thus, the seller would be better off by limiting her market share or reducing the quality of the clientele data at its disposal. *However, note that the optimal strategy of the seller is not unique.* Any signal $\{\kappa_{s_i}\}$ that satisfies (8) would lead to the same expected profit. Nevertheless, irrespective of the strategy chosen by the seller, it's incentive to lower the precision of consumer data increases as the buying firms' actions become strong substitutes. As a matter of fact when $\beta \to -\infty$ - the case of perfect substitutes, no trade takes place at equilibrium. The seller's optimal strategy is to sell uninformative data with $\kappa_{s_i^*} \to 0$ to all buying firms at a price $p_i^* \to 0$.

Privacy Implications for Society - When $\beta \geq -(1+\frac{\kappa_z}{\kappa_x})$, the information trading market behaves like one of 'strategic

complements' (and this includes the pure complements case where $\beta > 0$) and thus without sufficient (regulatory) penalties (or enforcing a data minimization principle) on adverse effects of consumer data sale, there could be considerable risk due to information trading when sanitized versions of private consumer data is under sale. The privacy implications for the case when the information trading market behaves like one of 'strong substitutes', i.e., when $\beta < -(1 + \frac{\kappa_z}{\kappa_x})$ favor a comparatively better "privacy-friendly" society when compared to the complements case. This is simply because the sold data is distorted (and distortion increases with increasing λ) and devoid of moderate to high correlations and thus it is difficult, compared to the complements case, to breach consumer privacy.

IV. Market Analyses with Additional Features

We analyze in the light of [2], single seller, multiple buyer privacy trading markets with additional general features. More specifically we consider two problems. First, we analyze the state of privacy in society when single seller, multiple buyer privacy trading markets incorporate *purposeful* information exchanges (either via observing behavior of competing buyers, through other paid third party sources, or even via collusion and/or cooperative inference between firms) between strategic buying firms to infer the quality of consumer data floating in the market before it takes it action. Second, we account for costs borne by the seller of consumer data (e.g., the costs related to the disutility faced by consumers in the event of a privacy breach), and also the costs borne by the strategic buying firms (e.g., costs incurred to make sense of the data at the disposal of the firms for further processing and learning/mining analysis) in processing the data they buy/use.

A. Purposeful Data Collection on Competing Buying Firms

In this section, we analyze privacy trading markets where buying firms are consciously keen to get market information on their competitors' data quality. In practice such information can be obtained either via (i) observing selling price of competing buyers obtained individually or through other paid third party sources, (ii) collusion between firms[8] and/or (iii) third-party information about the data quality possessed by other buying firms. We assume that, in addition to prior x_i, and the bought data s_i, buying firm i can also condition its action on *seepage data*, S_i given by:

$$S_i = A + \nu_i,\ \nu_i \sim \mathcal{N}(0, \frac{1}{\kappa_\nu}), \tag{9}$$

where $A = \int_0^1 a_i di$ denotes the aggregate action, and ν_i are independently distributed across the buying firms. In practice, κ_ν acts as a proxy for the extent of information seepage in the market as an aggregate measure, about competing firms. S_i is perfectly informative about the aggregate action A when $\kappa_\nu = \infty$, however as κ_ν decreases, the preciseness of the information content of the seepage data is reduced. In the boundary case when $\kappa_\nu = 0$, data S_i does not contain any payoff-relevant information, and reduces to the no-seepage setting in our model. Recall here from Section II that the payoff function for each firm is such that it cares about the actions of other firms only insofar as these actions impact the aggregate action A. This implies that any setting in which buying firm i observes noisy data about other firms' individual actions can be mapped into an isomorphic setting in which firm i only observes a signal about the aggregate action. To formally model each buying firm's ability to account for consumer data in their strategic decisions, we use the framework introduced by Vives [14], and extend the buying firms' strategy space by appending S_i to the strategy function $a_i(\cdot, S_i)$ for each i that maps its private (prior) and market (bought) consumer related information, (x_i, s_i) to an action that depends on the seepage data S_i. Thus, the Bayesian Nash equilibrium of the subgame between the buying firms require (i) each firm to choose $a_i(x_i, s_i, S_i)$ in order to maximize its expected profit conditional on its information set, i.e., $\mathrm{E}[\pi_i | x_i, s_i, S_i]$, taking the strategies of all the other firms as given; and (ii) the aggregate action needs to be consistent with the realization of the buying firms' individual actions, i.e., $A = \int_0^1 a_i(x_i, s_i, S_i) di$. We have the following result regarding the outcome for the seller and the buying firms at market equilibria, the proof of which is in the Appendix [13].

Theorem 4.1: At the unique Bayesian Nash equilibrium of the buying oligopoly game, for sufficiently small $\kappa_\nu > 0$,

1) *The seller's profit decreases in the extent of data seepage, i.e., $\frac{\partial \Pi^*}{\partial \kappa_\nu} < 0$; and*
2) *There exists $-(1 + \frac{\kappa_z}{\kappa_x}) < \bar{\beta} < 0$ such that $\kappa_{s^*} < \kappa_z$ for all $\beta \in (-(1 + \frac{\kappa_z}{\kappa_x}), \bar{\beta})$.*

Theorem Intuition and Practical Implications - The theorem states that regardless of whether buying firm actions are strategic substitutes or complements, the seller's profits decrease as the degree of data seepage is intensified. This is due to the fact that a firm's willingness to pay for consumer data reduces because they could free-ride on the information purchased from the seller by competing firms. The higher the degrees of seepage between firms, the seller is increasingly incentivized to charge lower prices for the consumer data at her disposal, thereby making lesser profits. The result also establishes that the range of β's for which the seller of consumer information finds it optimal to distort that information, increases in the presence of increased seepage. Recall from Theorems 3.4 and 3.5 that with no information seepage, the seller would decrease the quality of consumer data *if and only if* $\beta < -(1 + \frac{\kappa_z}{\kappa_x})$. In contrast, part (b) of Theorem 4.1 states that, no matter how small the amount of seepage, the seller would offer distorted data for some $\beta > -(1 + \frac{\kappa_z}{\kappa_x})$. This is due to the fact that the seller's ability to extract surplus from the buying firms by increasing the precision of offered data s_i for firm i, is reduced in the presence of seepage - simply because firms can free-ride instead of buying from the seller. However, the fact that $\bar{\beta} < 0$ means that, regardless of the presence or absence of information seepage, distorting *to be sold* consumer data is not optimal when the buying firms' actions are strategic complements.

Privacy Implications for Society - In the case of data seepage

[8] Anti-trust laws will come into action here, and thus by 'collusion' we imply actions by buying firms that do not lead to lawsuits.

between firms, the satisfaction of $\beta \geq -(1+\frac{\kappa_z}{\kappa_x})$ does not ring warning bells for privacy (as it did previously) even though it includes the 'complements' setting - the free riding behavior of buying firms ensures distortion of the offered consumer data as an incentive compatible strategic action by the profit-minded seller. This, in the case without the presence of penalties. However, though there exists some $\beta \geq -(1+\frac{\kappa_z}{\kappa_x})$ for which free-riding ensures data distortion by the seller, thereby taking positive steps towards preventing privacy breaches, there does also exist a large enough set comprising β values in the range $[-(1+\frac{\kappa_z}{\kappa_x}), \infty]$ that entails the negative privacy effects of markets with strategic complements. Thus, on average we can expect purposeful data seepage between buying firms to be a more privacy-friendly setting (regardless of complements or substitutes), compared to no-seepage settings.

B. Accounting for Transaction and Processing Costs

So far, we have analyzed privacy trading markets in the absence of penalties in the event of facing adverse effects of trading consumer data. In this section, we model various costs that are accrued by both the supply and the demand side in a privacy trading market. More specifically, we consider costs accrued by (a) the data seller in releasing clientele data to buying firms, including *costs to address potential post transaction privacy breaches* that might occur due to the trade, and (b) the processing *costs that buying firms undertake with the data they buy from the seller.* We note here that apart from the costs to address privacy concerns, the data seller (mobile apps) can face, as part of transaction costs, unpopularity costs related to their clients experiencing delay and high cellular bandwidth costs in loading these sites.

From a data buyer perspective, we assume that the buying firms face a processing cost that is quadratic in the quantity of utility garnered by a firm through its action. Mathematically, we represent a buying firm i's profit accounting for the processing cost as follows:

$$\pi(a_i, A, \theta) = \alpha_0 v(a_i)u(\theta) + \alpha_1 v(a_i)v(A) - \frac{1}{2}c_i v(a_i)^2, \quad (10)$$

where $c_i > 0$ is the cost accrued by the buying firm per unit of utility, A is the aggregate action in the market taken by the buying firms, and $\alpha_0, \alpha_1 < 0$ are constants. In practice, the processing costs are incurred by the data buyers to make sense of the data at the disposal of the firms for further processing and learning/mining analysis in processing the data they buy/use. Note the subtle difference between Equations (1) and (10) - the addition of the c_i parameter per unit of utility in the cost function in (10). In (1), we had assumed a unit cost for c_i homogeneously across all buyers indicating the same technology for data processing. In (10), we relax this assumption to model the more realistic case of relative (w.r.t. 1) heterogeneous costs to process bought data using varying technologies that are deployed by the firms.

We assume that the data seller incurs a transaction cost that equals $v \cdot \kappa_{s_i}$ whenever it sells consumer data of precision κ_{s_i}, where v is the constant (for the purpose of simplicity) cost per unit of precision offered by the data seller to the buyers. In practice, this transaction cost can imply four cost aspects: (i) the cost to offer verifiable and credible (appropriately distorted) consumer data to the buyers, (ii) the cost to cover premiums payable to cyber-insurance providers in the event of data breaches occur and consumers hold the data seller liable, (iii) the cost to customize consumer information to meet buyer needs, and (iv) facing consumer dissatisfaction with respect to slow and bandwidth consuming ad-embedded seller app. We have the following result, based upon [2], regarding the case involving transaction and processing costs, the proof of which is in the Appendix [13].

Theorem 4.2: At the non-unique Bayesian Nash equilibrium of the buying oligopoly game, there exist $\bar{v} > \underline{v}$ such that

1) *if $v > \bar{v}$, the data seller does not find it incentive compatible to trade with any downstream buying firm, i.e., $\kappa^*_{s_i} = 0$, for all i.*
2) *if $v < \underline{v}$, the seller finds it incentive compatible to offer selling consumer data without any distortion to the entire population of buying firms.*
3) *for any $v \in (\underline{v}, \bar{v})$, there exists a c^* such that*

$$\kappa^*_{s_i} = \begin{cases} 0, \text{ if} c_i > c^* \\ \kappa_z, \text{ if} c_i < \frac{\kappa_x^2}{(\kappa_x+\kappa_z)^2}c^*. \\ \kappa_x(\sqrt{\frac{c^*}{c_i}} - 1), \text{ otherwise} \end{cases}$$

Theorem Intuition and Practical Implications - We first note that v is the cost borne by the seller per-unit of precision offered to the buyers. Thus, higher the precision, i.e., quality, more the transaction costs. The theorem states that regardless of whether buying firm actions are strategic substitutes or complements, the seller finds its optimal to offer its buyers consumer data at a precision level that is decreasing in cost c_i - implying that firms having lesser processing cost per utility of consumer data get offered higher quality data. In addition, the seller's offered precision $\kappa^*_{s_i}$ at a market equilibrium is always non-increasing in c_i, for any buying firm i. This however does not imply that for a buying firm accruing a processing cost of zero, the seller is going to sell the latter an undistorted version of consumer data. More specifically, $\underline{v}$ reduces with reduced c_i and as long as the former is non-negative and per-unit precision cost $v > \underline{v}$ to the seller less than $\bar{v}$, the seller finds it incentive compatible to sell increased precision signals with decreased c_i, for any buying firm i - the boundary case arising if $v \leq \underline{v}$ in which case the seller finds it incentive compatible to sell distortion-free data. The intuition here is that the interval $(\underline{v}, \bar{v})$ denotes the zone where the seller gets enough business revenue via quality data (not necessarily undistorted) sold, at the same time does not need to sell enough contracts so as to incur costs of privacy risks characterized by v. This situation can arise for consumer data driven services that do not run high enough privacy risks (e.g., entertainment apps). However, if the costs of privacy are high enough, i.e., $v > \bar{v}$, the seller does not sell at all. This situation can arise for services characterized by highly privacy-sensitive data (e.g., medical and health data) that can be estimated relatively easily by downstream buyers from each other or through the environment. Furthermore, $\underline{v} < 0$ implies

$$\int_0^1 \frac{1}{c_i} di < -\frac{1}{\alpha_1}(1 + -\frac{\kappa_z}{\kappa_x}),$$

which is identical to the condition of Theorem 3.4 that did considered equal costs incurred by the buyers.

Privacy Implications for Society - The privacy implications for this theorem are quite similar to those of Theorems 3.4 and 3.5, overall. Thus, even in the presence of transaction and processing costs accrued by the seller and buyers, respectively, strategic complement settings increase privacy risks in society, where substitutes settings reduce privacy risks. The differences lie in the thresholds at which the risks aggravate or alleviate. These thresholds are driven by the parameters in Theorem 4.2 - those that consequently drive the seller to offer and sell consumer data at various quality and quantities.

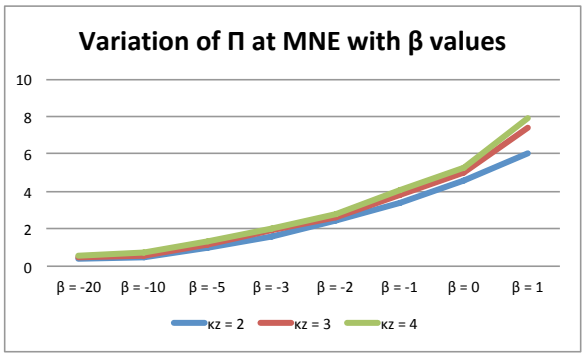

Fig. 2: Seller profit (units) variation with β at the MNE

V. Trace-Driven Markets Evaluation

In this section, we conduct a trace-driven market evaluation of our model. We numerically study the tradeoffs at market equilibrium between the profit earned by a seller, the data quality being channeled downstream to the buyers, and information privacy risks posed in society. In addition, we study, using a standard gradient descent approach, the convergence speed our oligopolistic market model to market equilibria, in a distributed fashion.

A. Trace-Driven Real-World Evaluation Setup

The Environment Setting - We collect name-sanitized (to preserve anonymity) consumer data for 1000 clients on their two sleep patterns (i.e., time to go to sleep, hours of sleep) from a fitness app startup firm (representing the data seller) based in northern California, USA. For policy considerations, the data for the clients are scaled proportionally by the startup firm before our collection to prevent us from experimenting with real personal data. For the aggregate data collected from the company, we set up three independent (of the firm) sleep expert representatives, A, B, and C from a medical department at an university in northern California to *act* as competing ad-networks. The experts have at least ten years of experience in research and consulting, and more importantly possesses deep knowledge of what type of sleep data would be of interest to different commercial organizations (representing advertisers in the supply chain) in the fitness and pharmaceutical industries. Having collected real-world data, as a mock experiment, we synthetically implement a triopoly downstream competition between A, B, and C as buyers of sanitized client data.

The Parameter Settings - Without loss of generality, we assume γ and α_2 to take a value of 1. In this regard, note that no matter what the value of α_2, α_0 and α_1 can be scaled accordingly. A similar logic follows the selection of a γ value for the purpose of simulation. We fix κ_x (the precision parameter of the prior of consumer data visible to a buying firm) equal to 1 and vary κ_z (the precision parameter of consumer data at the disposal of the selling firm) to take values of 2, 3, and 4. In addition, we vary the degree of strategic substitutability, β from 0 (non-strategic) to -20 (a high degree of substitutability) in intervals. Note here that the modulus of the negative β values would reflect the strategic complement property of sold data. For the case when purposeful information transfer related to consumer data happens between the buying firms, we vary κ_ν (the proxy parameter for the degree of information leakage) from 0 (no leakage) to 10 (high leakage) in intervals. The scalar utility value to a buying firm for a given instance of a consumer data matrix offered by the seller is derived using a linear map, i.e., the trace of the matrix, to a scalar value.

Evaluation Objectives (EOs) - We evaluate the following objectives in this paper for information trading market settings:

1) The variation of κ_s^* (offered data quality) and λ^* (quantity of contracts offered) at Nash equilibria with β.
2) The variation of the market Nash equilibria profit of the seller with the intensity of market competition (governed by β) between buying firms.
3) The variation of the seller profit ratio (ratio of profit with and without data distortion) at market Nash equilibria with β.
4) The variation of κ_s^* and λ^* at market Nash equilibria with β, at different degrees of purposeful consumer data related information transfer between the buying firms.
5) The variation of the market Nash equilibria profit of the seller with the intensity of competition governed by β, at different degrees of purposeful information transfer between the buying firms.
6) The pairwise correlation of consumer data quality bought by the buying firms at market Nash equilibria, reflecting the extent of information privacy breach risks.
7) The convergence speed to market equilibria, in terms of the number of iterations, of our market mechanisms.

B. Evaluation Analysis

We observe from Figure 3, that plots EO1, that the offered data quality at market equilibrium, denoted via κ_s^* increases with a reduction in the degree of strategic substitutes (denoted by β) and converges to κ_z at threshold values of β. A similar trend can be associated with the amount of contracts offered by the data seller at a market Nash equilibrium - the entire population being offered at threshold values of β. In addition, with increased κ_z values, i.e., reduced precision of the base quality acquired by the data seller, at a market Nash equilibrium (MNE) κ_s and λ values increase relative to those where κ_z values are lower. The rationale for this trend being that the seller finds it economically incentive compatible to offer relatively higher quality consumer data for sale to an increased fraction of buyer population, as its acquired precision levels of clientele data drop. Even as a

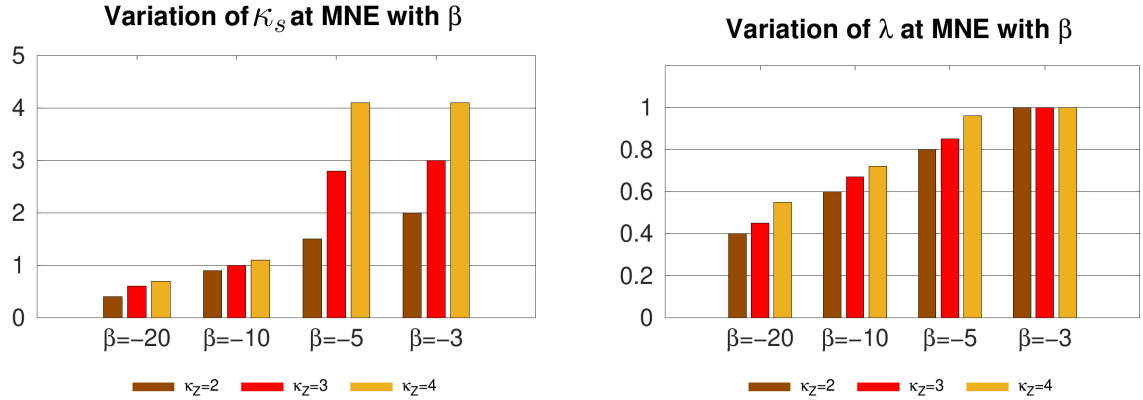

Fig. 3: Seller offered quality (left) and quantity (right) variation with β at Market Nash Equilibrium (MNE)

behavioral justification this makes sense as one (the seller) would want to make up for market reputation reasons, its lack of availability of high quality client data. With respect to EO2, we observe from Figure 2 that the MNE profit (in units of profit) of the seller increase (upto approximately 20 fold based on our model) with an increase in β - also an indicator of market competition. This trend arises as with increase in market competition, the seller opens up on its selling range and quality of data offered to cover the entire buying base. In addition, similar to EO1, with increased κ_z values, at an MNE, seller profit values increase - precisely for the same reason as that in EO1.

Figure 5 plots trends for EO3 where we compare the ratio of seller profit when it distorts data before selling in contrast to the situation when it does not - both with respect to the non-strategic case, i.e., when $\beta = 0$. It is evident that with zero strategic substitutes nature of the data sold, the ratio is 1 (for $\frac{\Pi_{\beta=0}^{d/nd}}{\Pi_0}$'s) as there is no data distortion for $\beta \geq 0$. For β values less than 0 (the case of less market competition and increasing substitutes nature of offered data), the ratio decreases due to lesser population being offered a sale at MNE, and that too at reduced data quality. The increase in profit between the $\frac{\Pi_{\beta=0}^{d}}{\Pi_0}$ and $\frac{\Pi_{\beta=0}^{nd}}{\Pi_0}$ cases being higher when the market competition is lower and substitutes property of offered data is higher - in our case approximately leading to a 100% improvement when $\beta = -20$. Moreover, similar to that in EO1 and EO2, with increased κ_z values, at an MNE, seller profit ratio values increase - for the same reason at those in EO1 and EO2.

In Figures 6 and 7, we plot trends pertaining to EO4 and EO5, respectively. We observe that with increase in the degree of information transfer between buying firms, indicated via κ_ν values, κ_s, λ, and Π values decrease with increasing κ_ν. The rationale for this trend being that transfer of information does not make it economically incentive compatible for the selling to improve the quality and quantity (hence profit) of offered clientele data downstream to competing buyers - the extent of the drop approximately being (model based) upto 100% (in κ_s), 40% (in λ), and 100% (in profit Π) at an MNE.

Figure 8 pertains to EO6 and represents the pairwise correlation, $C(\eta_i, \eta_j)$, in the data offered by the seller to the downstream buying firms. This correlation measure (equivalent to mutual information (MI) [15][5]) is a reflection of the information privacy (IP) risks in an information theoretic sense [16], posed in society due to trading personal information of consumers - *higher the correlation, greater the risk of an IP breach between competing firms and personal data percolating deep down in the supply chain without the intent or consent of the consumer base.* We plot $C(\eta_i, \eta_j) = \frac{\kappa_\xi + \rho_\xi \kappa_z}{\kappa_\xi + \kappa_z}$ - the pairwise correlation measure (see Section II) at the MNE for different ranges of β values. Here ρ_ξ is computed using the standard distance correlation (*dcor*) correlation metric [17] that accounts for non-linear dependencies between random variables. We observe that as β decreases, so does the value of $C(\cdot, \cdot)$ at MNE and that too at a rapid rate (that of increasing marginals) - indicating much reduced risks of IP breach events on decreased market competition and increased substitutes nature of traded data. The correlation value is 1 when $\beta \geq 0$, because highest quality data is offered to the entire population of buyers, and subsequently pose the greatest risk to IP.

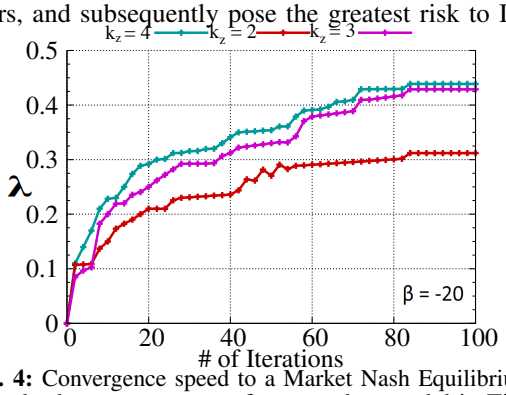

Fig. 4: Convergence speed to a Market Nash Equilibrium

We study the convergence of our market model in Figure 4, as part of EO7. To this end, we use a standard gradient descent algorithm [18] on our real-world induced data, that converges in a distributed fashion. Our primary performance metric is convergence speed in terms of the number of iterations. In Figure 4, we plot the evolution of λ to its market equilibrium value when $\beta = -20$ for different κ_z values. We note here that this is a just a representative example of the multiple parameters, i.e., the κ_s's and Π's, that converge to an MNE. Other examples report similar trends. We observe that for almost all examples, markets converge to a Nash equilibrium within 30 iterations of a Macbook Pro laptop (2017 version) with 16GB RAM. This at least indicates the possibility of the existence of markets if personal data were to be traded in a small-sized market consisting a few buyers. *As part of future plans, we plan to run larger scale field experiments to validate our claims on information markets of large buyer sizes.*

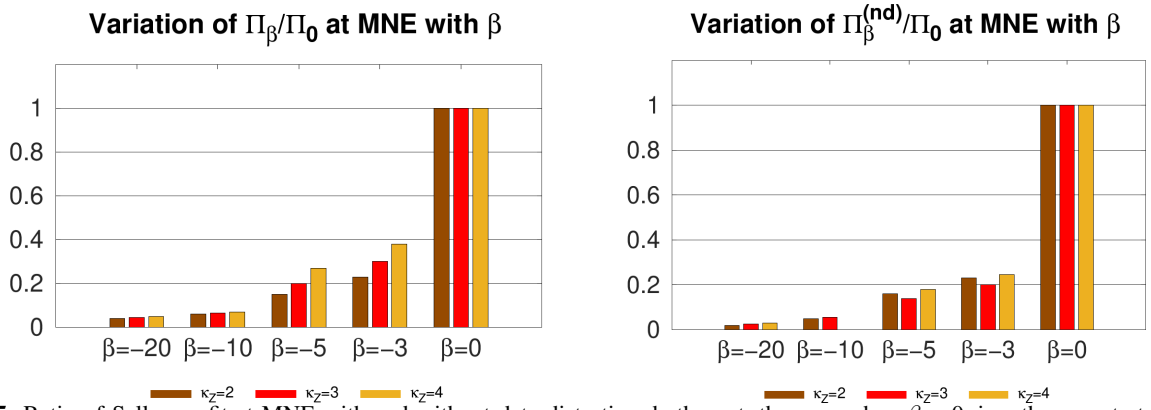

Fig. 5: Ratio of Seller profit at MNE with and without data distortion, both w.r.t. the case when $\beta = 0$, i.e., the non-strategic case

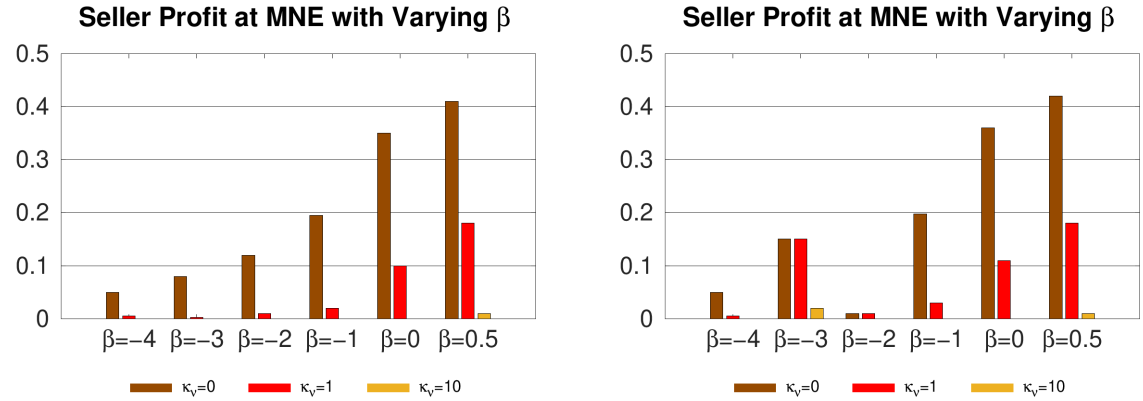

Fig. 6: Seller profit at an MNE with degrees of information transfer between buying firms, when $\kappa_z = 1$ (left) and $\kappa_z = 2$ (right)

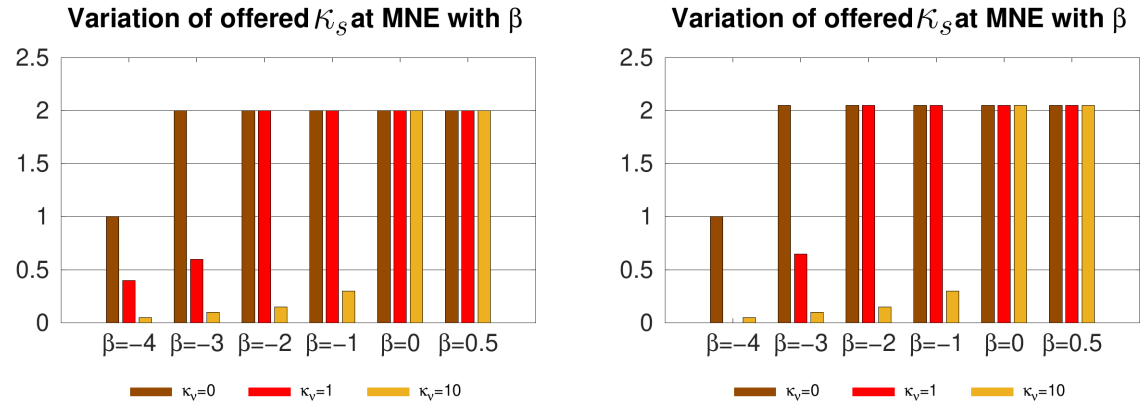

Fig. 7: Ratio of Seller profit at MNE with and without data distortion, both w.r.t. the case when $\beta = 0$, i.e., the non-strategic case

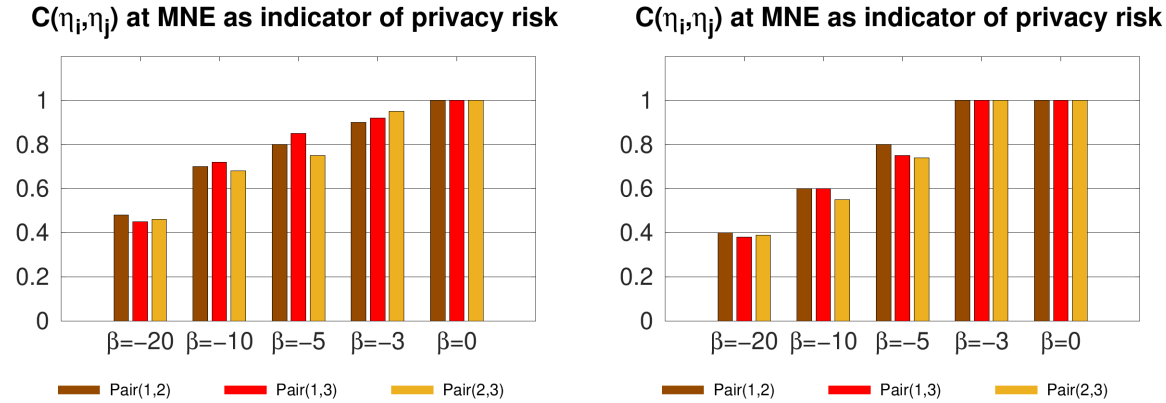

Fig. 8: $C(\eta_i, \eta_j)$ values at MNE as indicator of privacy breach risk for different pairs of buyers, when $\kappa_z = 2$ (left) and $\kappa_z = 3$ (right)

VI. Related Work

Research in relation to mapping the economic type of information with the degree of privacy risks they pose, is new to the best of our knowledge. In this section, we briefly review related literature pertaining to the existence and design of information trading markets. We identified two strands of research in this context: one rooted in the economics literature, and the other rooted in the technical literature on privacy-aware mechanism design.

The vision and benefits for information (privacy) trading had their roots in arguments made in the 1970s by Chicago school economists, Posner[19][20] and Stigler[21], in favor of having increased social welfare. In later years, their arguments were upvoted by information economists such as Laudon[22] and Acquisiti[23] Varian [24], Odlyzko [25], Schwarz [26], and Samuelson [27]. The primary thesis of these scholars being that the lack of use of personal client data will lead to opportunity costs and market inefficiencies (sub-optimal states of economic social welfare) since it conceals potentially relevant information from other economic agents (the downstream data intermediary entities in Figure 4) that eventually hamper the profitability of these agents. In contrast to the Chicago-school views, a number of economists including Hirshleifer[28][29], Burke[30], Wagman[31],Daughety & Reinganum[32], and Spence [33] are of the opinion that the costs to the demand side of the market to acquire quality client information may outweigh its social benefit, thereby decreasing social welfare. According to Varian[24], Odlyzko[25], and Acquisiti[23], consumer data obtained (with or without consent) can have negative effects on society welfare simply because post transaction the consumers have little knowledge or control over how and by whom their personal data will later be used. This conjecture has recently been formally proven by [34][35]. The firm (e.g., ad-networks) may sell the consumer's data to third parties (e.g., advertisers), which may lead to spam and adverse price discrimination, among other concerns, and subsequently lead to consensual consumers opting out of trade in future. Regulation here can curb the adverse effects of these negative externalities arising from trading and significantly contribute to welfare efficient and complete markets (where supply equals demand) [36][37]. Examples of practical ways to implement regulations suggested in existing literature include legislative property rights on consumer personal data shared between the supply and demand side[22], and technical metrics (e.g., differential privacy) being adopted by demand side data intermediaries such as ad-networks to keep a check on the degree of privacy breach[38]. From a non-privacy perspective, De Corni et.al. [39] state that targeted advertising in markets driven by the presence of private *non-perturbed* consumer information can lead to higher equilibrium prices. This result is in line with Levin and Milgrom[40], Bergemann and Bonatti[41], and Cowan[42] who design markets improving match quality by disclosing consumer information to downstream firms. The common takeaway message from these papers is that such markets might be too costly to a data intermediary — because of the informational rent that is passed on to selling firms.

Most existing technical works on privacy-aware mechanism design [43][44][45][46][47][48][49] assume that there is a trusted data holder (e.g., app). The private data, i.e., personal consumer information, is either already kept safely to itself by the data holder, or is evoked using mechanisms that are designed with the aim of consensual truthfulness - i.e., What the data holder purchases is the "right" of using consumers' data in an announced way. A major direction in which our work differs from existing work is in considering that data holders are not trusted by consumers to keep their data private, and may release it to agencies like ad-networks in return for benefits. To this end, in the seminal trading work by [43], consumers' data is already known to the data collector (the data collector here analogous to a buying side ad-network in our work), and the selling side (analogous to the data holder in our work) bid their costs of privacy loss caused by data usage by the buying side, where each seller's privacy cost is modeled as a linear function of ϵ if its sold data is used in an ϵ-differentially private manner. The goal of the mechanism design here is to evoke truthful bids of seller cost functions. In contrast, our setting is (a) more realistic in terms of data seller cost functions that are assumed to be not linear but convex, and (b) orthogonal to the [43] setting in the sense that we deal with a single seller trading with multiple buyers, rather than multiple sellers and a single buyer as in [43] - the latter popularly calling for an auction-mechanism approach. In addition, we also investigated on what types of information trading in supply-chain data release frameworks, conditioned on the economic type of sold data, will pose less privacy risks in society, something of primal interest to regulators, and not addressed by any existing work in literature. As a practical use-case of modeling privacy trading, the authors in [50] design a privacy trading mechanism for commercializing location privacy in mobile crowdsensing services. More specifically, they propose an auction-theoretic framework between workers and the platform to trade location privacy data, given a differential privacy induced leakage budget. However, though they are similar in nature to our motivation in trading privacy leakage, there is a significant fundamental difference between their contribution and ours: we formally model oligopolistic market competition between established buyer firms being served by a data seller; in contrast, the players (workers) in [50] are mobile end users distributed in a geographical locality thereby only interacting with the platform through an auction, and not traditionally competing in an oligopoly market.

Subsequent works [44][45][46][48] explore various models for seller valuation of privacy, especially the correlation between the cost functions and the private bits. This line of work has been extended to the scenario that the data is not available yet and needs to be reported by the sellers to the data collector, but the data collector is still trusted [47][51][52][49] - whereas we assume that the data collector (the ad-network in our case) is purposely buying consumer data from DHs (apps) for selling to advertisers in return for monetary gains from the latter. For more details on the interplay between differential privacy and mechanism design, [53] gives a comprehensive survey. In [54], the authors envisage a market model for private data analytics such that private data is treated as a

commodity and traded in the market. In particular, the buyer (the ad-network in our case) uses a game-theoretic incentive mechanism to pay (or reward) sellers for reporting informative data, and the sellers control their own consumer data privacy by reporting noisy data with the appropriate level of privacy protection (or level of noise added) being strategically chosen to maximize their payoffs. Like us, the authors in [54] assume that utility parameters of individuals (buyers in our case) are not private information. However, unlike them or as in any of the above-mentioned works, we deal with the (orthogonal) more practical problem of managing *heterogeneous* privacy guarantee demands between the data seller (e.g., apps) and multiple individual buyers (e.g., ad-networks), whereas the above works deal with homogeneous privacy guarantees. Very recently, the authors in [55] address the heterogeneous privacy guarantee case. However, to address information asymmetry on the seller side, their solution is restricted to the design of a two-seller, single buyer contract based on a binary distribution of seller privacy attitudes. In contrast, our solution is general and addresses the multi-seller, single buyer setting, where seller preferences in information asymmetry scenarios are captured using supply functions.

VII. Summary

In this paper our main goal was to investigate whether specific consumer data types have a varying influence on information privacy risks. The answer to this question will guide policy makers to regulate information flow for ad-exchange driven data markets in ways to maintain privacy interests of upstream consumers. To this end, we investigated and analyzed single-seller, multiple buyer information trading markets where a consumer facing data seller (e.g., mobile app) sells its aggregate clientele data in a privacy-sanitized manner downstream to multiple strategic buyers (e.g., ad-networks, retailers). We analyzed trading markets, built upon recent work in [2], distinguished by two mutually exclusive economic types of data being traded that entails different strategic actions types by the downstream buyers: (i) those exhibiting strategic substitutes and (ii) those exhibiting strategic complements. As our main result, we showed that information trading markets dealing with strategic substitutes data generally pose less information privacy risks for society at market equilibrium, compared to markets dealing with strategic complements data. This is primarily because (a) buyers are not economically incentivized to buy substitutes data from a seller when they could (robustly) estimate such data, via a free-riding mechanism or otherwise, from a correlation analysis of competing buyer information, and (b) the seller does not gain in profit to sell substitutes data above a certain threshold fraction of strategic buyers and in addition is incentivized to add extra noise to the data - thereby reducing IP risks arising from statistical correlations. We observed that our main results hold for both, markets where the seller faces a significant penalty costs for privacy breach events, and liberal markets where no costs are accrued by the user in the event of a privacy breach of consumer data. We also analyzed the state of privacy in society when single seller, multiple buyer information trading markets incorporate *purposeful* privacy-sanitized data exchanges (either via free-riding, observing behavior of competing buyers, collusion, or through other paid sources) between strategic buying firms. *Under public knowledge of this assumption*, our analysis resulted in similar findings as mentioned above - mainly that a profit-minded data seller is incentivized economically to (a) not sell data to all possible buyers, and (b) sell more noisy consumer data to the chosen buyers compared to the noise in the case when there is assumed to be no data exchange between buying firms. As a result, risk of privacy breaches are reduced in society. Our contributions in this paper will serve as a recommendation tale to regulators to effectively use a tradeoff knob when it comes to allow trading appropriate quantity and quality of economic types of data by selling firms.

VIII. Acknowledgement

This work is supported by the NSF under grants CNS-1616575, CNS-1939006, CNS-2012001, and ARO W911NF1810208.

References

[1] R. Pal and J. Crowcroft, "Privacy trading in the surveillance capitalism age viewpoints on'privacy-preserving'societal value creation," *ACM SIGCOMM Computer Communication Review*, vol. 49, no. 3, pp. 26–31, 2019.
[2] K. Bimpikis, D. Crapis, and A. Tahbaz-Salehi, "Information sale and competition," *Management Science*, vol. 65, no. 6, pp. 2646–2664, 2019.
[3] S. Mehta, M. Dawande, G. Janakiraman, and V. Mookerjee, "How to sell a dataset? pricing policies for data monetization," in *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 679–679, 2019.
[4] A. Agarwal, M. Dahleh, and T. Sarkar, "A marketplace for data: An algorithmic solution," in *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 701–726, 2019.
[5] A. Rényi, "On measures of dependence," *Acta mathematica hungarica*, vol. 10, no. 3-4, pp. 441–451, 1959.
[6] E. A. Haggard, "Intraclass correlation and the analysis of variance.," 1958.
[7] A. Mas-Colell, M. D. Whinston, J. R. Green, *et al.*, *Microeconomic theory*, vol. 1. Oxford university press New York, 1995.
[8] R. Pal, J. Crowcroft, Y. Wang, Y. Li, S. De, S. Tarkoma, M. Liu, B. Nag, A. Kumar, and P. Hui, "Preference-based privacy markets," *IEEE Access*, vol. 8, pp. 146006–146026, 2020.
[9] P. D. Klemperer and M. A. Meyer, "Supply function equilibria in oligopoly under uncertainty," *Econometrica: Journal of the Econometric Society*, vol. 57, no. 6, pp. 1243–1277, 1989.
[10] D. Fudenberg and J. Tirole, "Game theory," 1991.
[11] I. L. Glicksberg, "A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points," *Proceedings of the American Mathematical Society*, vol. 3, no. 1, pp. 170–174, 1952.
[12] P. Dasgupta and E. Maskin, "The existence of equilibrium in discontinuous economic games, i: Theory," *The Review of economic studies*, vol. 53, no. 1, pp. 1–26, 1986.
[13] "Appendix of privacy risk is a function of information type," Google Docs. https://drive.google.com/file/d/1uW63JpjMCej˙UsVeQ8fs8yOpOgd0KRgB/view.
[14] X. Vives, "Strategic supply function competition with private information," *Econometrica*, vol. 79, no. 6, pp. 1919–1966, 2011.
[15] R. Pal, P. Hui, and V. Prasanna, "Privacy engineering for the smart micro-grid," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 5, pp. 965–980, 2018.
[16] I. Wagner and D. Eckhoff, "Technical Privacy Metrics: A Systematic Survey," *ACM Computing Surveys (CSUR)*, vol. 51, 2018.
[17] G. J. Székely, M. L. Rizzo, *et al.*, "Brownian distance covariance," *The annals of applied statistics*, vol. 3, no. 4, pp. 1236–1265, 2009.
[18] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*, vol. 23. Prentice hall Englewood Cliffs, NJ, 1989.
[19] R. Poser, "The right of privacy," *Georgia Law Review*, vol. 12, no. 3, 1978.
[20] R. Poser, "The economics of privacy," *American Economic Review*, vol. 71, no. 2.
[21] G. Stigler, "An introduction to privacy in economics and politics," *Journal of Legal Studies*, vol. 9, no. 4, 1978.
[22] K. C. Laudon, "Markets and privacy," *Commun. ACM*, vol. 39, pp. 92–104, Sept. 1996.
[23] A. Acquisti, C. Taylor, and L. Wagman, "The economics of privacy," *Journal of Economic Literature*, vol. 54, no. 2, pp. 442–92, 2016.

[24] H. Varian, "Economics aspects of personal privacy," *Privacy and Self-Regulation in the Information Age*, 1997.

[25] A. Odlyzko, "Privacy, economics, and price discrimination on the internet," *Economics of Internet Security (Eds. Jean Camp, Stephen Lewis*, 2003.

[26] P. M. Schwartz, "Property, privacy, and personal data," *Harv. L. Rev.*, vol. 117, p. 2056, 2003.

[27] P. Samuelson, "Privacy as intellectual property?," *Stanford law review*, pp. 1125–1173, 2000.

[28] J. Hirschleifer, "The private and social value of information and the reward to inventive activity," *American Economic Review*, vol. 61, no. 4, 1971.

[29] J. Hirschleifer, "Privacy: Its origin, function, and future," *Journal of Legal Studies*, vol. 9, no. 4, 1980.

[30] J. Burke, C. Taylor, and L. Wagman, "Information acquisition in competitive markets: An application to the us mortgage market," *American Economic Journal: Microeconomics*, vol. 4, no. 4, 2012.

[31] L. Wagman, "Good news or bad news?: Information acquisition and applicant screening in competitive labor markets," *SSRN*, 2014.

[32] A. Daughety and J. Reinganum, "Public goods, social pressure, and the choice between privacy and publicity," *American Economics Journal: Microeconomics*, vol. 2, no. 2, 2010.

[33] M. Spence, "Job market signalling," *Quarterly Journal of Economics*, vol. 2, no. 2, 2010.

[34] R. Pal, J. Li, Y. Wang, M. Liu, S. De, and J. Crowcroft, "Data trading with a monopoly social network: Outcomes are mostly privacy welfare damaging," *IEEE Networking Letters*, 2020.

[35] R. Pal, Y. Wang, J. Li, J. Crowcroft, M. Liu, S. Tarkoma, and Y. Li, "Data trading with a competitive social platforms: Outcomes are mostly privacy welfare damaging," *IEEE Transactions on Network and Service Management*, 2020.

[36] R. H. Coase, "The problem of social cost," in *Classic papers in natural resource economics*, pp. 87–137, Springer, 1960.

[37] P. Bolton and M. Dewatripont, *Contract Theory*. MIT Press, 2005.

[38] R. Pal and J. Crowcroft, "Privacy trading in the age of surveillance capitalism: Viewpoints on 'privacy-preserving' societal value creation," *ACM SIGCOMM Computer Communication Review*, vol. 49, no. 3, 2019.

[39] A. de Corniere and R. De Nijs, "Online advertising and privacy," *SSRN*, 2014.

[40] J. Levin and P. Milgrom, "Online advertising: Heterogeneity and conflation in market design," *American Economic Review*, vol. 100, no. 2, 2010.

[41] D. Bergemann and A. Bonatti, "Targeting in advertising markets: Implications for offline versus online media," *RAND Journal of Economics*, vol. 42, no. 3, 2011.

[42] S. Cowan, "The welfare effects of third-degree price discrimination with non-linear demand functions," *RAND Journal of Economics*, vol. 38, no. 2, 2007.

[43] A. Ghosh and A. Roth, "Selling privacy at auction," *Games and Economic Behavior*, vol. 91, pp. 334–346, 2015.

[44] L. K. Fleischer and Y.-H. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 568–585, ACM, 2012.

[45] K. Ligett and A. Roth, "Take it or leave it: Running a survey when privacy comes at a cost," in *International Workshop on Internet and Network Economics*, pp. 378–391, Springer, 2012.

[46] A. Roth and G. Schoenebeck, "Conducting truthful surveys, cheaply," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 826–843, ACM, 2012.

[47] A. Ghosh and K. Ligett, "Privacy and coordination: computing on databases with endogenous participation," in *Proceedings of the fourteenth ACM conference on Electronic commerce*, pp. 543–560, ACM, 2013.

[48] K. Nissim, S. Vadhan, and D. Xiao, "Redrawing the boundaries on purchasing data from privacy-sensitive individuals," in *Proceedings of the 5th conference on Innovations in theoretical computer science*, pp. 411–422, ACM, 2014.

[49] A. Ghosh, K. Ligett, A. Roth, and G. Schoenebeck, "Buying private data without verification," in *Proceedings of the fifteenth ACM conference on Economics and computation*, pp. 931–948, ACM, 2014.

[50] W. Jin, M. Xiao, M. Li, and L. Guo, "If you do not care about it, sell it: Trading location privacy in mobile crowd sensing," in *IEEE INFOCOM*, IEEE, 2019.

[51] D. Xiao, "Is privacy compatible with truthfulness?," in *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pp. 67–86, ACM, 2013.

[52] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan, "Truthful mechanisms for agents that value privacy," *ACM Transactions on Economics and Computation (TEAC)*, vol. 4, no. 3, p. 13, 2016.

[53] M. M. Pai and A. Roth, "Privacy and mechanism design," *ACM SIGecom Exchanges*, vol. 12, no. 1, pp. 8–29, 2013.

[54] W. Wang, L. Ying, and J. Zhang, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 44, pp. 249–260, ACM, 2016.

[55] M. M. Khalili, X. Zhang, and M. Liu, "Contract design for purchasing private data using a biased differentially private algorithm," in *Proceedings of the 14th Workshop on the Economics of Networks, Systems and Computation*, p. 4, ACM, 2019.

Ranjan Pal is a faculty member of ECE at University of Michigan Ann Arbor. His primary research interest lies in engineering robust cyber-security and information privacy solutions using tools from decision and the applied mathematical sciences. Ranjan received his PhD in Computer Science from USC's Viterbi School of Engineering, and was a postdoctoral fellow at the University of Cambridge and USC. He is a member of IEEE, ACM, American Mathematical Society, INFORMS, SIAM, and Game Theory Society.

Junhui Li is working toward dual bachelor's degree in both computer science at the University of Michigan Ann Arbor and electronic and computer engineering in Shanghai Jiao Tong University. Her primary research interests lie in machine learning, privacy and security in systems, and human computer interaction. She is a student member of the IEEE.

Jon Crowcroft is the Marconi Professor of Communications Systems in the Computer Laboratory at the University of Cambridge. His current active research areas are Opportunistic Communications, Social Networks, Privacy Preserving Analytics, and techniques and algorithms to scale infrastructure-free mobile systems. Since 2016, he has been Programme Chair at the Alan Turing Institute. He is a Fellow the Royal Society, a Fellow of the ACM, a Fellow of the British Computer Society, a Fellow of the IET and the Royal Academy of Engineering and a Fellow of the IEEE.

Yong Li (M'09) received the B.S. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007, and the Ph.D. degree in electronics engineering from Tsinghua University, Beijing, China, in 2012. In 2012 and 2013, he was a Visiting Research Associate with Telekom Innovation Laboratories and The Hong Kong University of Science and Technology, respectively. From 2013 to 2014, he was a Visiting Scientist with the University of Miami. He is currently a Faculty Member with the Department of Electronic Engineering, Tsinghua University. His research interests are in the areas of networking and communications.

Mingyan Liu is an entrepreneur and the Peter and Evelyn and Fuss Chair Professor of Electrical and Computer Engineering at University of Michigan Ann Arbor. She received her PhD in Electrical Engineering from University of Maryland College Park. Her current research interests lie in communication networks, sequential decision theory, incentive design for cybersecurity and privacy, online learning, and experimental data science related to cyber security. She was a co-founder of the cybersecurity scoring startup Quadmetrics in 2014 that got acquired by FICO in 2016. She is a Fellow of the IEEE and a member of the ACM.

Nishanth Sastry is a Senior Lecturer (Associate Professor) at King's College London, UK. He holds a PhD in Computer Science from the University of Cambridge. His current research interests include computer and social networks, computational social science, and data analytics aspects of these two areas. He has been a visiting researcher at the Alan Turing Institute and Massachusetts Institute of Technology. His research has been granted nine patents in the USA for work done at IBM. Nishanth is a member of the ACM and a Senior Member of the IEEE.