IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 27, 2020, accepted July 27, 2020, date of publication August 7, 2020, date of current version August 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3014882

Preference-Based Privacy Markets

RANJAN PAL[1,2,8], (Member, IEEE), JON CROWCROFT[2], (Fellow, IEEE), YIXUAN WANG[1], (Student Member, IEEE), YONG LI[3], (Senior Member, IEEE), SWADES DE[4], (Senior Member, IEEE), SASU TARKOMA[5], (Senior Member, IEEE), MINGYAN LIU[1], (Fellow, IEEE), BODHIBRATA NAG[6], (Senior Member, IEEE), ABHISHEK KUMAR[5], (Student Member, IEEE), AND PAN HUI[5,7], (Fellow, IEEE)

[1]Department of Electrical Engineering and Computer Science (EECS), University of Michigan, Ann Arbor, MI 48109, USA
[2]Department of Computer Science and Technology, University of Cambridge, Cambridge CB2 1TN, U.K.
[3]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[4]Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India
[5]Department of Computer Science, University of Helsinki, 00100 Helsinki, Finland
[6]Indian Institute of Management Calcutta, Kolkata 700104, India
[7]Computer Science and Engineering Department, Hong Kong University of Science and Technology, Hong Kong
[8]Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB2 1TN, U.K.

Corresponding author: Ranjan Pal (palr@umich.edu)

This work was supported in part by the NSF under Grant CNS-1616575 and Grant CNS-1939006, in part by the Army Research Office (ARO) under Grant W911NF1810208, in part by the Hong Kong Research Grants Council under Project 16214817, in part by the 5GEAR Project, Academy of Finland, under Grant 319017, and in part by the FIT Project from the Academy of Finland.

ABSTRACT In the modern era of the mobile apps (the era of *surveillance capitalism* - as termed by Shoshana Zuboff) huge quantities of surveillance data about consumers and their activities offer a wave of opportunities for economic and societal value creation. In-app advertising - a multi-billion dollar industry, is an essential part of the current digital ecosystem driven by free mobile applications, where the ecosystem entities usually comprise consumer apps, their clients (consumers), ad-networks, and advertisers. Sensitive consumer information is often being sold downstream in this ecosystem 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 breaches of very significant impact (e.g., breach of genetic data) in the short term. The question we raise through this paper is: *Is it economically feasible to trade consumer personal information with their formal consent (permission) and in return provide them incentives (monetary or otherwise)?*. In view of (a) the behavioral assumption that humans are 'compromising' beings and have privacy preferences, (b) privacy as a good not having strict boundaries, and (c) the practical inevitability of inappropriate data leakage by data holders downstream in the data-release supply-chain, we propose a design of regulated efficient/bounded inefficient economic mechanisms for oligopoly data trading markets using a novel preference function bidding approach on a simplified sellers-broker market. Our methodology preserves the *heterogeneous* privacy preservation constraints (at a grouped consumer, i.e., app, level) upto certain compromise levels, and at the same time satisfies information demand (via the broker) of agencies (e.g., advertising organizations) that collect client data for the purpose of targeted behavioral advertising.

INDEX TERMS Information privacy, preference, supply function economics, trading, market equilibrium.

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. As a popular example, *Evite.com* may sell lists of their consumers attending a party in a given location to advertisers via ad-networks run by Google and Facebook. Similarly, the gene testing company *23andMe* might sell their clientele information directly to pharmaceutical companies in order for the latter to develop medical drugs. As a social

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

objective, a 'win-win' deal between (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. The vision and benefits for such an ecosystem were laid down by a certain school of information economists way back from the 70's (see more details in [1]), in favor of having increased aggregate societal welfare. More specifically, according to the survey, in return for personal data, advertisers and marketers will benefit the consumer side through monetary compensation (e.g., discounts, Facebook Libre coins) and intangible benefits (e.g., personalization and customization of information content), and price discrimination. Furthermore, the same school of information economists state that the lack of use of personal data might lead to opportunity costs and market inefficiencies. To furthermore emphasize the benefits of privacy trading, now from a consumer viewpoint, a survey conducted by the authors in [2] advocate consumers willing to trade data for incentives. *In this paper, we take the side of these economists to investigate privacy outcomes in society as a result of such markets.* However, before we lay down research contributions with respect to such markets, we provide an explanation of why such markets are a need of the day despite privacy concerns raised due to IP commercialization.

A. NEED FOR FAIR PRIVACY COMMERCIALIZATION

Most would agree that doing business with consumer data without their consent is outright creepy. Consequently, as a landmark regulatory corrective step to prevent commercialization of personal data, the General Data Protection Regulation (GDPR) was initiated in May 2018 that impose constraints, rights, obligations, and voluntary consumer choice regarding personal data and its use. However, it is questionable as to whether the psychological approach of many apps—in offering a binary voluntary opt in/out, often after presenting pages of legalese—results in user empowerment with respect to making the proper choice between gaining utility from an app versus not using it. Indeed, we see that individuals are increasingly using ad-blocking technology[1] as a means to 'push-back', alongside deciding to gain utility from apps. However, ad blocking firms like *Eyeo*, maker of the popular *AdBlock Plus* product, has achieved such a position of leverage that it gets Google et.al., to pay it to have their ads whitelisted by default - under its self-styled 'acceptable ads' program [3] - clearly going against the of the core functionality principle of ad-blockers.

Thereby, with a significant likelihood, there might be an inevitable breach of personal consumer information in general to satisfy the economics behind the working of the current ad ecosystem. According to a recent study [3] conducted post GDPR enactment, influential popular app-firms like New York Times (NYT) can likely make more revenues from traditional advertising channels such as TV/newspapers, compared to online/mobile advertising. However, this argument might not hold for moderate sized firms who consequently would rely heavily on behavioral advertising for generating revenues. The bottomline here is data intermediary entities will commercially gain from the consumer data release downstream, whereas psychologically tricked consumers, some of them being under the effect of the privacy paradox [4], voluntarily give up their personal data and lose out on both privacy and monetary gains - *an unfair proposition.* Moreover, one could argue here that paying for apps[2] would mitigate this issue, however, statistics prove that consumers around the world are more keen on using free apps compared to paid apps,[3] and are also quite neutral to the collection of cookies by third parties, during browsing activities.[4]

On an orthogonal (to regulatory issues) note, Shoshana Zuboff in her recent book [5] states with numerous real-life surveillance examples of how since the early 2000's (primarily after 9/11), our daily life activities and 'deepest secrets' are all recorded, rendered as behavioral data, processed, analysed, bought, bundled, and resold like sub-prime mortgages in a *behavioral futures* market, thanks to companies such as Google and Facebook whose initial motivations for data collection were rooted in boosting ROI for their investors. And in seeking to survive commercially beyond their initial goals, these companies realised they were sitting on a new kind of asset: our 'behavioural surplus', the totality of information about our every thought, word and deed, which could be traded for profit (via rejecting established norms of societal responsibility and accountability) in new markets based on predicting, shaping, and controlling our every need - or producing it. The extraction of such information assets by tech giants is so grotesque, so creepy, that it is almost impossible to see how anyone who really thinks about it lives with it - and yet we do. There is something about its opacity, its insidiousness, that makes it hard to think about. Likewise the benefits of faster search results and turn-by-turn directions mask the deeper, destructive predations of what Shoshana Zuboff terms 'surveillance capitalism', a force that is as

[1] https://pagefair.com/blog/2017/adblockreport/

[2] There are quite a few services that already offer some level of choice/configuration between full subscription (no ads, thus no third party privacy exposure) and fully advertisement/analytics paid for (i.e. "free"). Consequently there's the possibility of doing an empirical study to populate a model of peoples'(not yet evident that they are privacy-rational) "willingness to pay" in terms of utility function/curves for privacy/money.

[3] https://www.appsflyer.com/resources/state-app-spending-global-benchmarks-data-study/

[4] Statistic.com

profoundly undemocratic as it is exploitative, yet remains poorly understood - a central strategy of this regime. Despite more and more people expressing their unease about the surveillance economy, and seeking alternatives, it might be long before we extricate ourselves from the toxic products of both industrial and surveillance capitalism. *Till then, one workable solution might be to trade consumer data with their consent in a fashion that benefits all fairly in the data release ecosystem, and not just the data greedy firms.* To this end, the reader is referred to our recently published work, [6], for additional details on the rationale behind privacy trading being a solution jointly aligned with the supply and demand sides of a privacy market.

B. TOWARDS 'Preference-Based' TRADING

A deeper look into existing research in the generic area of designing privacy preserving economic mechanisms (courtesy the survey paper in [7], though the paper is not in line with the idea of privacy trading as applicable to this work) reveals that the fundamental inability for *any* economic mechanism dealing with consumer data to achieve a social optimal state with respect to privacy (be it for data trading ecosystems or otherwise) lie in (i) the hardness to satisfy *strict* heterogeneous consumer privacy preferences, and (ii) the inability to internalize the negative externalities due to privacy leakage, e.g., recent Facebook-Cambridge Analytica data scandal [8]. Thus, *as our main idea, a direction towards optimizing social welfare, i.e., economic efficiency, is to relax the strictness of privacy preserving preferences*, thereby *allowing heterogeneous consumers to compromise their ideal privacy requirements with their permission/consent in return for benefits* (e.g., monetary and non-monetary incentives). These benefits contribute to resolving the issue in (ii).

The weight behind this novel idea of ours lies in the fact that from a psychological perspective, most human beings are acceptable to making varied levels of compromises in real-life, especially for goods like privacy that have non-clear boundaries [2] (See Section VIII for few examples where privacy compromises are acceptable). Note that privacy compromises by consumers would result in apps selling more relevant personalized information to ad-networks (and thereby generating more revenue), the latter able to sell more ad-space to advertisers at an increased revenue, and the advertisers being able to target a broader personalized set of consumers. Thus, we have a win-win situation among all ecosystem entities. The big question then is: *what is an optimal way to compromise aggregate consumer privacy?*

Research Goal - As a major goal, we aim to investigate via a theory methodology, our radical idea of optimally compromising aggregate consumer privacy, in a simplified market ecosystem, through the combined use of micro-economic theory and a composition property characteristic of the family of information-theoretic privacy preserving technologies. Here, the term 'optimal' is in the sense of achieving maximum utilitarian social welfare as an economic efficient state. Through our efforts, we wish to provide introductory foundational insights on designing information trading markets that improve social welfare, and pave the way for a more general analysis of complex trading markets.

C. RESEARCH CONTRIBUTIONS

We make the following research contributions in this paper.

- We model a privacy trading ecosystem setting as a supply-demand market consisting of (i) market competing (both, in perfect and oligopolistic fashion) data holders (DHs) representing app firms with locked-in consumer base and (ii) a single ad-network acting as a data broker between the app firms and the advertisers. A salient feature of this trading ecosystem is the use of data holder *supply functions* [9] - privacy preference functions that map the amount of privacy compromise (the 'supply') at an aggregate consumer level each data holder is willing to make, i.e., the supply, *for* a given "benefit" it receives from the ad-network per unit of data. The data holders submit their supply functions as bids to an ad-network that then executes a uniform market clearing "benefit" mechanism for all competing data holders, to achieve optimal utilitarian privacy welfare at market equilibria (see Section III).[5]
- We analyze perfectly competitive (in DHs) and oligopolistic privacy trading markets based on our proposed supply function model, for existence, uniqueness, and economic efficiency of market equilibria. For perfectly competitive markets we show that they achieve a maximum utilitarian social welfare state, i.e., an economic efficient state, at a unique equilibrium. However, for oligoplistic trading markets, we show that they reach a unique market equilibrium that does not maximize utilitarian privacy welfare in society (see Section IV).
- We mathematically characterize the efficiency loss for oligopolistic trading markets by quantifying the difference between the unique market equilibrium obtained in the competitive scenario with that in the oligopoly scenario, via a Price of Anarchy (PoA) measure. As major results, we find the following: (a) the set of data-holders at oligopolistic Nash equilibrium (ONE) who compromise on their privacy requirements at the aggregate consumer level, is a superset of that at the perfectly competitive equilibrium (PCE); (b) the market clearing "benefit" (per unit of compromise) at the ONE is higher than that at the PCE, but the ratio of the two "benefits" is bounded; (c) the sum total of data holder disutility (e.g., due to privacy compromise of their clients) at ONE is larger than that at PCE, but the ratio is bounded by certain mild assumptions; (d) if data holders have relatively homogeneous cost functions (e.g., for trading data types with similar privacy sensitivities), the differences between the PCE and ONE tend to be very small - if the

[5]The readers are referred to the Section VIII (due to space constraints) for a qualitative introduction on supply function economics and its relevance to this work

cost functions are extremely heterogeneous (for trading data types with different privacy sensitivities), the quantification of the differences can serve as rules of thumb for the ad-network to limit the privacy compromising power of DH firms to promote utilitarian social welfare. For each of (a)-(d), we provide practical implications pertaining to privacy and policy. (see Section V).

- We show in Section V that for the problem at hand, our proposed supply function mechanism for privacy trading is *optimal* over a feasible family of mechanisms.

II. RELATED LITERATURE

In this section, we briefly review related literature most relevant to privacy 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. With respect to privacy-preserving metrics of operation, applicable only to the technical literature, we note that the metric proposed in this work is assumed to fall in the same general family of metrics used in existing works, i.e., the family of information-theoretic privacy (IP) metrics (see [10]) where resulting data is encapsulated with generated statistical noise to preserve IP, and IP guarantees are additive (e.g., as in differential privacy (DP)).

The vision and benefits for information (privacy) trading (not necessarily consensual) had their roots in arguments made in the 1970s by University of Chicago economists, Posner [11], [12] and Stigler [13], in favor of having increased social welfare. In later years, their arguments were upvoted by information economists such as Laudon [14] and Acquisiti *et al.* [1] Varian [15], Odlyzko [16], Schwarz [17], and Samuelson [18]. 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 (e.g., the downstream data intermediary entities in Figure 1) that eventually hamper the profitability of these agents. As a modern day example, client data (obtained via apps) on fitness, health habits, cyber-hygiene can benefit (cyber) insurance service agencies to target and allocate well-matched policies to their clients - conversely the lack of quality data can lead to bad matches and erode profit margins. In contrast to the Chicago-school views, a number of economists including Hirshleifer [19], [20], Burke *et al.* [21], Wagman [22], Daughety and Reinganum [23], and Spence [24] are of the opinion that the costs to the demand side of the market to acquire quality client information in a non-consensual setting may outweigh its social benefit, thereby decreasing social welfare. It is here that consensual information trading with benefits to the supply side could reduce the costs to acquire supply side information and improve social welfare. *In this work, we adopt the Chicago school of thought and assume that sellers will be consensual with the buyer demands in return for monetary remuneration.*

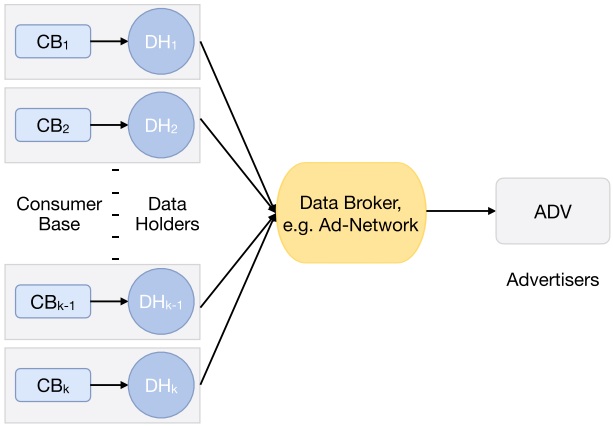


FIGURE 1. Market Architecture with a Single Data Broker (Ad-Network).

We assume consensual information trading to be regulated in the interest of social welfare, and an appropriate step for determining the effectiveness of trading in data intermediary settings such as in Figure 1. According to Varian [15], Odlyzko [16], and Acquisiti *et al.* [1], consumer data obtained (with or without consent) can have negative effects on society simply because post transaction the consumers have little knowledge or control over how and by whom their personal data will later be used. 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) [25], [26]. 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 [14], technical metrics (e.g., DP) being adopted by demand side data intermediaries (e.g., ad-networks) to check on the degree of IP breach [6], and frameworks such as those developed in [27]–[30] to improve security and privacy for BigData systems (e.g., HDFS).

Specifically, in relation to the data intermediary settings such as in Figure 1, De Corni'ere and Nijs [31] rule out, for regulated consensual trading settings, direct price discrimination by the demand side on the supply side based on consumers' personal information by focusing instead on advertising firms' bidding strategies in auctions for more precise targeting of their advertisements. That is, given that consumers' private information provides a finer and finer segmentation of the population, firms can compete to advertise their non-discriminatory pricing over each of those consumer segments. Hence, by disclosing information about consumers, the ecosystem ensures that consumers will see the most relevant advertisements, whereas when no information is disclosed under a complete privacy regime, ads are

displayed randomly. *This is in contrast to our model that vouches for price-discrimination - the reason being in our setting, unlike the above-mentioned works, there is a statistical perturbation of the consumer private data sold downstream with noise for privacy considerations. Hence a finer clear segmentation is not possible.* De Corni et.al. also state that targeted advertising in the presence of private non-perturbed consumer information can lead to higher prices, and, in line with Levin and Milgrom [32], Bergemann and Bonatti [33], and Cowan [34] that improving match quality by disclosing consumer information to firms might be too costly to an intermediary - because of the informational rent that is passed on to selling firms. *This is again in contrast to our findings - simply because in our model the selling data might be perturbed downstream by statistical noise.*

Most existing works on privacy-aware mechanism design [35], [36] [37], [38] [39], [40], [41] assume that there is a trusted data holder of unperturbed consumer data. The private data is either already kept by the data holder, noise perturbed by it, or is evoked using mechanisms that are designed with the aim of truthfulness. What the data holder purchases is the ''right'' of using individuals' 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, may not noise perturb it to appropriate levels while releasing it to agencies like ad-networks, in return for benefits.* To this end, in the seminal work by [35], individuals' data is already known to the data collector (the data collector here analogous to an ad-network in our work), and individuals (analogous to the data holder in our work) bid their costs of privacy loss caused by data usage, where each individual's privacy cost is modeled as a linear function of ϵ if his data is used in an ϵ-differentially private manner. The goal of the mechanism design here is to evoke truthful bids of individual cost functions. *In contrast, our setting is more realistic and assume that (a) DH cost functions are private information - not for release to an ad-network, and (b) cost functions need not be linear but convex.*

Subsequent works [36], [37] [38], [40] explore various models for individuals' (analogous to DHs in our work) 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 individuals to the data collector, but the data collector is still trusted [39], [42] [41], [43] - *whereas we assume that the data collector (the ad-network in our case) is purposely selling consumer data (obtained via DHs) to advertisers for monetary gains.* For more details on the interplay between differential privacy and mechanism design, [7] gives a comprehensive survey. In [44], 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 data collector (the ad-network in our case) uses a game-theoretic incentive mechanism to pay (or reward) individuals (DHs in our work) for reporting informative data, and individuals control their own 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. *However, unlike us, they assume that utility parameters of individuals are not private information, which may not be true in practice. In addition none of the above-mentioned works deal with the case of managing heterogeneous privacy guarantees across individuals (DHs in this work), as we do.* Very recently, the authors in [45] 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 are captured using supply functions.

In a very recent research effort, similar to our motivation, the authors in [46] design a privacy trading mechanism for commercializing location privacy in mobile crowdsensing applications. 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 work in proportionalizing benefits with privacy leakage (and showing budget-balanced, truthful, and incentive compatibility properties of auction mechanisms), there are some significant differences between the contributions made in [46] and this work: (i) we formally model market competition between established app firms serving a base of consumers; in contrast, the players (workers) in [46] 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 - hence such a market analysis is missing from their work, (ii) unlike us, the work in [46] neither characterize market efficiency gaps in theory, nor do they prove the optimality of their mechanism over feasible families of economic variables (e.g., cost functions, mechanism classes, etc.), and (iii) as an obvious distinction, our application space, i.e., a supply-chain framework of mobile apps leaking data upstream to ad-networks and advertisers, is different in geographical scope from that of mobile crowdsensing.

III. SYSTEM MODEL

In this section, we propose the salient features of our parameterized static market model representing a privacy trading ecosystem that is built atop the seminal economic theory of supply function bidding proposed by Klemperer in [9]. Due to space constraints, we refer the reader to a qualitative background (see [6]) of supply function theory by Klemperer and Meyer as being an appropriate *regulated* economic method that forms the primary basis in the design of markets to trade *group privacy*[6] - the privacy of a group of app clients, rather

[6]Shoshana Zuboff in her recent book, *The Age of Surveillance Capitalism* [5], states that it is group privacy that is most important to surveillance capitalists as the individual user is just a pawn and not the product - the product is group data.

than individual clients themselves. Table 1 can be referred to for a set of important notations used in the paper.

TABLE 1. Table of Important Notations.

$N, n = \|N\|$	set and number of data holders, i.e., DHs
q_i	privacy compromised amount for DH i
p_i	per unit of compromise benefit of DH i
b	bidding parameter
b^*	Nash equilibrium bidding profile
C_i	cost function for DH i
u_i	utility function of DH i
d	privacy compromise threshold
ONE	oligopolistic Nash equilibrium
PCE	perfectly competitive equilibrium
S_i	privacy compromise amount, DH i willing to take
π_i	payoff for DH i

A. MARKET ELEMENTS

Our market elements (see Figure 1) comprise of consumers locked in with their respective **data holders** (DHs) and an **ad-network** acting as a data broker between the data holders and a body of advertisers (ADV). We assume the presence of regulatory bodies (e.g., governments) whose goal is to ensure a certain level of social welfare state (e.g., maximum amount) keeping in mind the privacy interests of people in society.

We assume that consumers are locked-in with their respective data holders in a given time period. Examples of data holders include ad-publishing mobile apps, social media apps, IoT databox apps,[7] etc. Data holders compete with each other - as an example, competing mobile apps with similar functionalities (e.g., UberEats, GrubHub) are market competitors. Similarly, IoT databoxes manufactured by competing firms, each having their consumer base, compete with each other in the market. A consumer can simultaneously be client to multiple DHs. Based on pre-ordained policies, the data holders collect consumer data relevant to their functionality, and upon the consent of the consumers (e.g., Android and iOS phones have their own but different policies on how consumers can control data release to apps running on the phones). However, despite providing control to consumers, unwanted but voluntary data release by the latter is possible via methods designed through the proper use of psychology, behavioral economics, and neuroscience. Ad-networks (e.g., *Google Ad Network*, *Bing Ads* by Microsoft) act as mediators between DHs and advertisers, where the latter's goal is to post advertisements with DHs in order to enable targeting, tracking, and reporting of consumer impressions. Finally, to cite an example of the structure of data that could be traded by the DHs having access to aggregate consumer data from their client base - parts of it that is assumed to be private, a database is one of the possibilities. As popular practical examples, the firm *BookYourData* (BYD) offers upstream 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, job function, department, country etc.

[7] a given customer base can be associated with multiple competing app or social media DHs; however, in this work we assume a one-one mapping between consumers and DHs for relative tractable simplicity, as this setting itself is challenging enough. We leave the analysis of the one-many setting for future work.

B. MARKET STRUCTURE

We consider two traditional market structures: *perfect competition*, and *oligopoly*, to be operative amongst the DHs. In each structure, the competing DHs trade privacy compromise amounts with a single ad-network[8] using a supply function bidding[9] approach (see Section III.C). The ad-network in return provides some "benefits" (to be explained later in this section) to the DHs based on the amount of compromise made by the DHs. The ADV generates a demand[10] for consumer information to the ad-network, and in pay the ad-network to match them with appropriate DHs so as to enable targeting, tracking, and reporting of consumer impressions.

C. MODEL FOR SUPPLY FUNCTION BIDDING

In this section we formally introduce the mechanism between competing DHs and the ad-network. A diagrammatic illustration of the process as shown in Figure 2.

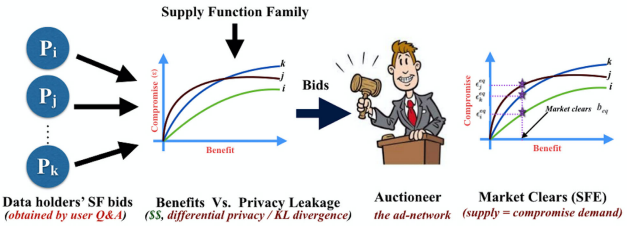


FIGURE 2. Illustrating Privacy Preference Function Trading with One Broker.

Setup - Consider a set N of $|N|$ DHs that are locked-in with their respective consumer base. In the ideal state, each DH needs to obey certain privacy requirements derived from the privacy preferences of their consumer base. To preserve generality, we assume that the privacy requirements of each DH map to a privacy metric that is an element of the set of *information gain metrics* [10] that measure the amount of information an adversary can gain. Note that the differential

[8] Since different ad-networks run their own supply function mechanisms for privacy trading independently of the others, the analysis of one extends to the others. Thus, each app will trade on different parameters with different ad-networks at market equilibrium (see Figure 3). Hence, in a somewhat simplistic sense, it is enough to analyse a single ad-network scenario. Moreover, when it comes to the number of major ad-networks, recent studies [3] report that they are primarily owned by *Google* and *Facebook*.

[9] Supply side privacy preferences, as functions of incentives, derived via survey Q&A, deviates us from the use of the standard Bertrand and Cournot trading mechanisms that have one-dimensional (price or quantity) strategy spaces.

[10] This is usually done through a bidding process like Vickrey-Clarke-Groves (VCG) auction (not the explicit focus of this work) between the ADVs and the ad-network, based on consumer data that interests relevant ADVs.

privacy metric is just one element of this set. Higher the value of the privacy metric, the less information an adversary can gain. However, given the presence of the ad-network and ADVs, *there are two main reasons why there may not be the simultaneous satisfaction of privacy requirements of each DH*: (i) keeping in mind the "benefit" making mindset of DHs (the "benefit" whose source are the ADVs), achieving the optimal cost-benefit tradeoff with the ad-network might not guarantee strict privacy-preservation for DHs, (ii) it is known, via results from [7], that designing mechanisms that ensure heterogeneous privacy preservation at a utilitarian social welfare optimal state is an open problem.

The Process - Each DH $i \in N$ is willing to consensually compromise $q_i(b_i, p_i)$ amounts of aggregate client privacy (measured through the privacy metric - as shown in Figure 2, usually either DP, KL-divergence, Mutual Information, etc.) with the ad-network, in return for a per-unit of compromise benefit value, p_i, i.e., q_i is a parameterized function of p_i and a non-negative bidding parameter b_i. As an example, let q_i to be a linear function of the form:

$$q_i(b_i, p_i) = b_i p_i, \quad i \in N, \tag{1}$$

The compromise function, q_i, for each DH i is their *parameterized supply function*. The benefit to each DH, p_i from the ad-network is primarily monetary in nature. Examples of benefits to the consumer base (derived from p_i[11]) include the amount of *price reduction* over the market price paid by individual consumers locked-in with a given DH[12] (for the case of paid apps), or in the case DHs are free to consumers - an amount of *reduction in the number of advertisements* displayed on the DH at a time instant (e.g., in case of an app) for each consumer to improve their usability experience.

We emphasize here that each DH i only submits the function q_i to the ad-network, as a signal of its preference on privacy compromise, without revealing its private utility/payoff function (see Section IV) of which q_i is just a part. Subsequently, the ad-network just has the values of q_i's at its disposal to arrive at a market uniform market clearing value of per-unit benefit that maximizes social welfare amongst the DHs[13]

We assume that the total privacy compromise demand for the ad-network coming upstream from the advertisers end needs to meet a specific amount $d > 0$ (for a general information-theoretic privacy measure) [14] when it clears the market, i.e.,

$$\sum_i q_i(b_i, p) = \sum_i b_i p = d, \tag{2}$$

or

$$p(b) = \frac{d}{\sum_i b_i}. \tag{3}$$

Note here that Equation (2) holds due to the composability property of certain privacy metrics such as differential privacy [47], [48]. $b = (b_1, \ldots, b_N)$ is the supply function profile of the DHs. In the event when $\sum_i b_i = 0$, the ad-network will reject the bid.

IV. MARKETS ANALYSES

In this section, we analyze perfectly competitive and oligopolistic market structures of DH competition in the backdrop of a single ad-network. The strategy space for the DHs is the set of feasible parameter values for their supply functions. We assume no restrictions on DH compromise amounts and select the linear supply function as the preferred choice for the DHs. To this end, we first provide a strong rationale on our choice of supply function. We then proceed with the markets analyses.

Why Use a Linear Supply Function? - We answer this question by first stating that, unlike us, the seminal work in [9] does use a general function as the bidding strategy for the purpose of analysis. However, if our bidding action were to change from the linear form (represented by the single variable, b_i in our work) to a general form like in [9], the analysis of the strategic behavior of the DHs become much more complicated. To drive home this point, solving the general supply function equilibrium (SFE) (introduced in [9]) requires solving a set of differential equations. To the best of our knowledge, there are only *existence* results about the SFE while assuming the agents (DHs in our work) are symmetric (i.e., with the same cost function) or assuming there are only two asymmetric agents - *these assumptions are not practical in reality.* For practical applications, the asymmetric case is more interesting. On the positive side, the greatest advantage of using linear supply function over the general forms is the ability to handle asymmetric DHs when there are more than two DHs. Moreover, as we will show later in this section, (a) the linear supply function allows us to get a closed form characterization for the structure and efficiency of the market equilibria, which could be impossible to get if using the general supply function, and (b) *in the case of oligopoly markets, linear supply function induced markets minimize worst case efficiency loss for non-restricted compromise markets.* Thus, we lose no generality in working with linear supply functions as they would be incentive compatible for rational DHs to use (see Section V).

A. PERFECTLY COMPETITIVE MARKETS

In perfectly competitive markets, DHs are 'benefit taking'. Such markets arise when there are a plethora of DHs selling similar basic consumer information (e.g., users' preferences towards the items or products, language preference, time zone) that are mostly not very personal - so a standard common benefit value ensues. Given a benefit p, each

[11] DHs make up for the discounts through benefits from the ad-network.

[12] The consumer market prices charged by competing DHs might vary for each DH.

[13] One could argue that the popular Kelly's mechanism would also suffice to obtain social welfare optimality, but the latter mechanism is suitable only for one-dimensional bids, and not necessarily functions.

[14] In the special case when the privacy metric under consideration is differential privacy, the total compromise demand d is analogous to the quantity ϵ_d from Section III, where $\epsilon_d = 0$ denotes a situation of zero compromise.

DH i maximizes its net revenue given as:

$$\max_{b_i \geq 0} pq_i(b_i, p) - C_i(q_i(b_i, p)) \tag{4}$$

where the first term is the revenue of DH i when it compromises $q_i(b_i, p)$ amount of privacy at a benefit p per unit of compromise with a bidding parameter of b_i, and the second term is the total cost incurred to make the compromise. This cost can be interpreted as the sum of (a) the cost of making technical adjustments required to compromise privacy (e.g., technological/software costs of hosting ads by advertisers), (b) costs of handling consumer complaints/unpopularity, (c) brand/app switching with respect to degradation of quality of experience (QoE) arising from clients experiencing delay and high cellular bandwidth costs in loading apps.

Definition 1: A perfectly competitive equilibrium (PCE) for the privacy compromise system is defined as a tuple $\{(\bar{b}_i)_{i\in N}, \bar{p}\}$ *such that* $\bar{p}_i$ *is optimal in (4) for each DH i given the benefit* $\bar{p}$ *and* $\sum_i q_i(\bar{b}_i, \bar{p}) = d$.

The following result shows the existence and uniqueness of PCE, and it also shows the efficiency of the latter in maximizing utilitarian social welfare. The proof of the theorem is in the Section VIII.

Theorem 1: The PCE, $\{(\bar{b}_i)_{i\in N}, \bar{p}\}$*, for the privacy compromise system exists and is efficient, i.e.,* $(\bar{q}_i)_{i\in N} = (q_i(\bar{b}_i, \bar{p}_i))_{i\in N}$ *maximizes the utilitarian social welfare amongst the DHs expressed mathematically as follows:* $max_{q_i\geq 0} \sum_i -C_i(q_i)$*, subject to* $\sum_i q_i = d$*. If the cost function* $C_i(q_i)$ *is strictly convex, the PCE is unique.*

Theorem Implications - The theorem implies that there exists a pure (and unique, if DH cost functions are strictly convex) strategy PCE vector of DH privacy compromise amounts for all DHs at a particular homogeneous PCE benefit $\bar{p}$ set by the ad-network that meets the aggregate ad-network demand of d units of total privacy compromise, and maximizes utilitarian social welfare amongst the DHs. *In a nutshell, the theorem states that at market equilibrium efficient privacy trading is possible amongst heterogeneous DHs and an ad-network.*

Based on the above theorem, we can further study how the compromise cost function affects a DH's privacy compromise amount at PCE. For each DH i, we define the base privacy compromise marginal cost as $C_i^0 = C_i'(0^+)$. Without loss of generality, we assume that $C_1^0 \leq C_2^0 \leq \ldots\ldots\ldots \leq C_{|N|}^0$. For modeling convenience, we also introduce parameter $C_{|N|+1}^0$ and set its value to $C_n'(d)$. Thus, we have $C_1^0 \leq C_2^0 \leq \ldots\ldots\ldots \leq C_{|N|}^0 \leq C_{|N|+1}^0$. We have the following result on the privacy compromise characteristics of individual DHs, the proof of which is in the Section VIII.

Theorem 2: Let $\{(\bar{b}_i)_{i\in N}, \bar{p}\}$ *be a PCE and* $\bar{q}_i = q_i(\bar{b}_i, \bar{p})$ *be the corresponding privacy compromise amount by DH i. The set of DHs that embrace positive compromise amounts, i.e.,* $\{i : \bar{q}_i > 0\}$*, at the PCE is given by the set* $\bar{N} = \{1, 2, \ldots\ldots, \bar{n}\}$*, with an* $\bar{n}$ *that satisfies*

$$\sum_i^{\bar{n}} (C_i')^{-1}(C_{\bar{n}}^0) \leq d \leq \sum_i^{\bar{n}} (C_i')^{-1}(C_{\bar{n}+1}^0). \tag{5}$$

Moreover, benefit $\bar{p}$ *at the PCE satisfies*

$$C_{\bar{n}^0} \leq \bar{p} \leq C_{\bar{n}+1}^0, \tag{6}$$

for any $i \in \bar{N}$, $\bar{p} = C_i'(\bar{q}_i)$.

Theorem Implications - The theorem states that the PCE has a waterfilling structure - the base privacy compromise cost $C_i'(0)$ determines whether DH i compromises privacy or not. The higher the marginal cost at zero, the less likely the DHs will join the privacy compromise program, i.e., embrace a positive amount of compromise. Moreover, the DHs who join the privacy program at PCE bear the same marginal cost. The theorem also implies individual rationality is guaranteed at PCE, i.e., each DH in the privacy compromise program makes non-negative net revenue - we state this as the following corollary, the proof of which is in the Section VIII.

Corollary 1: Any DH who participated in the privacy compromise program receives non-negative net revenue at PCE, i.e., $\bar{p}\bar{q}_i - C_i'(\bar{q}_i) \geq 0$ *for all* $i \in \bar{N}$.

Market 'Win-Win' for Ecosystem Stakeholders - An efficient privacy trading market implies that (a) DHs are led to optimal tradeoffs on how much to compromise aggregate client privacy versus the per-unit compromise (monetary) benefit they get from the ad-network, (b) the ad-network satisfies the downstream demand from the advertisers on their informational requirement, (c) advertisers, through the ad-network can get get their ads placed to the right audience, and (d) consumers, via the monetary benefits received by DHs from the ad-network, either get to pay less for their services, or view fewer ads to improve the QoE. They also see useful targeted ads.

B. OLIGOPOLISTIC MARKETS

In oligopolistic competition markets, DHs are 'benefit anticipating', i.e., the DHs know that the benefit p is set according to (3) and behave strategically. Such markets arise when there are a few DHs in the market *strategically* competing with one another on specific types of consumer information that might be sensitive to the latter (e.g., location, device ID, genetic information). We denote the supply function for all DHs but i as $b_{-i} = (b_1, b_2, \ldots, b_{i-1}, b_{i+1}, \ldots.., b_{|N|})$ and write (b_i, b_{-i}) for the supply function profile b. Each DH i chooses b_i to maximize its own benefit $u_i(b_i, b_{-i})$ given others' bidding strategy b_{-i}

$$u_i(b_i, b_{-i}) = p(b)q_i(p(b), b_i) - C_i(q_i(p(b), b_i)), \tag{7}$$

that simplifies to

$$u_i(b_i, b_{-i}) = \frac{d^2 b_i}{(\sum_j b_j)^2} - C_i\left(\frac{db_i}{(\sum_j b_j)}\right).$$

Here, the second equality is obtained by substituting the market clearing benefit $p(b) = \frac{d}{\sum_i b_i}$ and the linear supply

bidding function $q_i(p(b), b_i) = b_i p(b)$ into the first equality. As a result functions $\{u_i(b_i, b_{-i})_{i\in N}$ define a privacy compromise game.

Definition 2: A supply function profile b^* is an oligopolistic Nash equilibrium (ONE) if for all DHs $i \in N$, we have

$$u_i(b_i^*, b_{-i}^*) \geq u_i(b_i, b_{-i}^*), \quad \forall b_i \geq 0.$$

In order to derive results regarding the existence and uniqueness characteristics of Nash equilibria in oligopoly markets, we first propose the following *three* lemmas (for investigating the existence and uniqueness of ONE), which are proved in the Section VIII.

Lemma 1: If b^* is an ONE of the privacy compromise game, then $\sum_{j\neq i} b_j^* > 0$ for any $i \in N$.

Lemma 1 also directly implies the following lemma, which we state without proof.

Lemma 2: If b^* is an ONE of the privacy compromise game, then at least two DHs have $b_i^* > 0$.

Lemma 3: If b^* is a Nash equilibrium of the privacy compromise game, then $b_i^* < B_{-i}^* = \sum_{j\neq i} b_j^*$ for any $i \in N$, and each DH will compromise an amount less than $\frac{d}{2}$ at the ONE, and no ONE exists when $|N| = 2$.

The proof of Lemma 3 is in Section VIII. We now turn to state the *first* of the two main results in this section.

Theorem 3: Assume that $|N| \geq 3$. The privacy compromise game has a unique ONE. The ONE solves the following convex optimization problem:

$\min_{0\leq q_i<\frac{d}{2}} \sum_i D_i(q_i)$ subject to $\sum_i q_i = d$,
where

$$D_i(q_i) = \left(1 + \frac{q_i}{d - 2q_i}\right) C_i(q_i) - \int_0^{q_i} \frac{d}{(d - 2x_i)^2} C_i(x_i) dx_i.$$

Theorem Implications - The theorem implies that there exists a pure and unique ONE strategy vector of DH privacy compromise amounts for all DHs at a particular homogeneous ONE benefit p^* set by the ad-network that meets the aggregate ad-network demand of d units of total privacy compromise, *but does not provide a guarantee on maximizing utilitarian social welfare amongst the DHs* (see Section V in the paper for a mathematical explanation). *In a nutshell, the theorem states that at an oligopolistic privacy trading market between heterogeneous DHs and an ad-network leads to an equilibrium state that is not economically efficient.* From the proof of the theorem in the Section VIII, it can be seen as reverse-engineering from ONE to a global optimization problem. Define $\Delta C_i(q_i) = \frac{q_i}{d} - 2q_i C_i(q_i) - \int_0^{q_i} \frac{d}{(d-2x_i)^2} C_i(x_i) dx_i$. Then $D_i(q_i) = C_i(q_i) + \Delta C_i(q_i)$. Thus, $\Delta C_i(q_i)$ can be interpreted as ''false information'' reported by the DHs to gain more benefit from privacy compromise by the ad-network, through strategic bidding. Note that $\Delta_i C_i(q_i) > 0$ for all $q_i \in [0, \frac{d}{2})$. $\Delta_i C_i(q_i)$ being greater than zero implies that all DHs fake a higher cost function in order to increase the benefit.

Not the Best 'Win-Win' for Ecosystem Stakeholders - A 'no-guarantee' on the efficiency of privacy trading oligopoly implies that DHs might not be able to strategize in a manner so as to converge upon optimal compromise-benefit tradeoffs, but the existence of a unique market equilibrium suggests stable strategizing by the former, i.e., a win-win state that is not the best one. This means that the DHs will fake high costs of compromise to get more benefits that will transfer more incentives to the consumer side at ONE, when compared to PCE. However on the flip side, the privacy compromise amounts at ONE will be higher (not something the DHs would prefer) based on the true compromise costs of the DHs. From a privacy perspective, this result is fairly intuitive as various price strategic mobile apps *sell data that are correlated among the apps, and this correlation negatively affects privacy preservation guarantees at the ad-exchange.* The ad-network and the advertisers are able to satisfy their objectives, as in the PCE.

Based on Theorem 3, similar to the case of perfectly competitive markets, we can further study how a cost function affects a DH's privacy compromise amount at ONE. For each DH i, we define the base privacy compromise marginal cost as $C_i^0 = C_i'(0^+)$. Without loss of generality, we assume that $C_1^0 \leq C_2^0 \leq \ldots\ldots \leq C_{|N|}^0$. Also notice that $C_i'(0^+) = D_i'(0^+)$. For modeling convenience, we also introduce parameter $C_{|N|+1}^0$ and set its value to $\max_i D_{|N|}'(\frac{d}{3})$. Thus, we have $C_1^0 \leq C_2^0 \leq \ldots\ldots \leq C_{|N|}^0 \leq C_{|N|+1}^0$. We now have the *second* important result (see Section VIII for a proof) for this section, on privacy compromise characteristics of DHs.

Theorem 4: Let $|N| > 3$, $\{(b_i^*)_{i\in N}\}$ be an ONE, $p^* = \frac{d}{\sum_i b_i^*}$ be the ONE benefit, and $q_i^* = b_i^* p^*$ be the corresponding privacy compromise amount by DH i. The set of DHs i that embrace positive compromise amounts, i.e., $\{i : q_i^* > 0\}$, at the ONE is given by the set $N^* = \{1, 2, \ldots\ldots, n^*\}$, with an n^* that satisfies

$$\sum_i^{n^*} (D_i')^{-1}(C_{n^*}^0) \leq d \leq \sum_i^{n^*} (D_i')^{-1}(C_{n^*+1}^0) \tag{8}$$

Moreover, benefit p^* at the ONE satisfies

$$C_{n^*}^0 \leq p^* \leq C_{n^*+1}^0, \tag{9}$$

for any $i \in N^*$, $p^* = D_i'(q_i^*)$.

Theorem Implications - The theorem states that the ONE has a waterfilling structure, and henceforth the implications are exactly the same as for Theorem 2. The theorem also implies individual rationality is guaranteed at ONE, i.e., each DH in the privacy compromise program makes non-negative net revenue - we state this as the following corollary, the proof of which is in the Section VIII.

Corollary 2: Any DH who participated in the privacy compromise program receives non-negative net revenue at ONE, i.e., $p^* q_i^* - C_i'(q_i^*) \geq 0$ for all $i \in N^*$.

V. EFFICIENCY AND OPTIMALITY ASPECTS

In this section, we characterize efficiency loss of oligopoly privacy trading markets and derive the optimality of our mechanism choice.

A. CHARACTERIZING EFFICIENCY LOSS AT ONE

We have shown that utilitarian social welfare is maximized at PCE, thereby making perfectly competitive markets efficient. In contrast, due to DHs' benefit-anticipating and strategic behavior, the ONE is expected to be less efficient. In this section, we investigate the efficiency loss at ONE for different degrees of heterogeneity among DH cost functions, and provide closed form characterization of the efficiency loss (if any). Here, we define the the efficiency loss as the ratio of the total disutility at PCE to the minimum total disutility, i.e., the ratio $\frac{C^*}{\bar{C}}$. *Thus, efficiency loss is equivalently the price of anarchy (PoA) [49].* To this end, we have the following main result post investigation.

Theorem 5: Let $\{(\bar{b}_i)_{i\in N}, \bar{p}\}$ *be a perfectly competitive equilibrium (PCE), and* p^* *be the corresponding benefit at the oligopolistic Nash equilibrium (ONE). We have the following*:

1) $\bar{N} \subseteq N^*$ *where* $\bar{N}$ *is the set of DHs who participate in the privacy compromise program at PCE, and* N^* *is the set of DHs who participate in the privacy compromise program at ONE.*
2) $\bar{p} \leq p^* \leq n - \frac{1}{n} - \frac{2M}{m\bar{p}}$, *where* $M = \max_{i\in N} C_i'(\frac{d}{n})$; $m = \min_{i\in N} C_i'(\frac{d}{n})$.
3) $\bar{C} \leq C^*$, *and if we assume that* $\bar{q}_{\max} = \max_i \bar{q}_i < \frac{d}{2}$, *then we have*

$$C^* \leq (1 + \frac{\bar{q}_{\max}}{d} - 2\bar{q}_{\max})\bar{C},$$

where $\bar{C} = \sum_i C_i(\bar{q}_i)$ *be the total social cost at PCE, and* $C^* = \sum_i C_i(q_i^*)$ *is the total social cost at ONE.*

Theorem Implications - The conditions in the theorem together imply the following:

- The set of DHs that contribute to the privacy compromise program at ONE is a superset (due to more DHs seeing an opportunity to make benefits by bidding strategically) of that at PCE (due to the non-strategic nature of the DHs at PCE).
- The benefit at the ONE is higher than that at PCE (due to strategic DH behavior at ONE), but the ratio between the two benefits are bounded. This last point makes sure that there are limits of DHs to exploiting the advantage of strategic behavior over non-strategic behavior.
- The total (aggregate) compromise cost at the ONE is higher than that at the PCE (due to strategic higher bidding, consequently more benefits, consequently unwanted additional privacy compromise), but the ratio between the two costs are bounded (incentivizing strategic higher bidding over non strategic bidding), provided no one compromises more than half of the total demand at the PCE (can be enforced via regulation).
- In addition, as long as no DH compromises more than $\frac{d}{3}$ at PCE, the efficiency loss $\frac{C^*}{\bar{C}}$ is bounded by $\frac{3}{2}$. This condition can be guaranteed if there are at least three DHs having comparably low compromise cost (e.g., big firms with a huge base of locked-in clients and/or firms trading non-sensitive data), compared to the others. The presence of closed form expressions for the efficiency loss may serve as a guideline to regulators for limiting the market power of some DHs (in the oligopoly setting) to maximize social welfare (e.g., by allowing the entry of new moderate/big DH app firms in the market to stiffen competition, and/or control types of data to be traded).

Moreover, from Theorems 2 and 4, we can derive the following special case result if the DHs have homogeneous costs, and the difference between the two market equilibria, i.e., PCE and ONE, are small. The proof of the result is in the Section VIII.

Corollary 3: On the condition that DHs have the same cost function, we have the following: 1. $p^* = n - \frac{1}{n} - 2\bar{p}$. *As* $n \to \infty$, $p^* \to \bar{p}$. *2.* $C^* = \bar{C}$. *As* $n \to \infty$, $C^* \to \bar{C}$.

The condition guarantees that when app firms facing similar cost structure (due to trading similar data type) are in competition, applying the supply function bidding scheme will lead to system efficiency irrespective of whether the market is perfectly competitive or oligopolistic.

Can the Efficiency Loss be Unbounded? - We show with an example that the efficiency loss in the worst case can be unbounded. Consider the case where there are three DHs with cost functions $C_1(q) = \frac{1}{2rcq^2}$, and $C_2(q) = C_3(q) = \frac{1}{2cq^2}$, where c and r are constant parameters. Using Theorem 2, we can calculate the PCE to be: $\bar{q}_1 = \frac{r}{r+2d}$, $\bar{q}_2 = \bar{q}_3 = \frac{1}{r+2d}$, and $\bar{p} = \frac{r}{r+2cd}$. Similarly, using Theorem 3, we get the ONE as: $q_1^* = \frac{-r+\sqrt{(16+9r)r}}{4(2+r)d}$, $q_2^* = q_3^* = \frac{8+5r-\sqrt{(16+9r)r}}{8(2+r)d}$, and $p^* = D - \frac{q_1^*}{D} - 2q_1^*q_1^*$. Now let $r \to \infty$ - for the PCE we then have $\bar{q}_1 \to d$, $\bar{q}_2, \bar{q}_3 \to 0$, $\bar{p} \to cd$, and total cost $\bar{C} \to 0$. For the ONE, we have $q_1^* \to \frac{d}{2}$, $q_2^*, q_3^* \to \frac{d}{4}$, $p^* \to \infty$, and the total cost $C^* \to \frac{cd^2}{4}$. Thus, $\frac{p^*}{\bar{p}} \to \infty$, and $\frac{C^*}{\bar{C}} \to \infty$.

Message for Regulators - We see that in a market with DHs having extremely heterogeneous cost functions, the efficiency loss at the ONE might be unbounded. Combining this fact with the implications of Corollary 3, regulators are advised to enable privacy trading by apps in segregated pools, with similar data types to be traded.

B. OPTIMALITY OF OUR MECHANISM CHOICE

We prove the optimality of our mechanism choice, i.e., a linear supply function mechanism, over a class of mechanisms that are suited to designing markets for our problem.

To embark on this task, we first consider a mechanism desirable if it minimizes worst case efficiency loss when DHs are 'benefit anticipating', independent of the utility functions of the DHs and their number. That is, the mechanisms we seek are those that perform well under broad assumptions of the nature of the preferences of the market participants. We will show that under a specific set of assumptions, our mechanism choice minimizes the worst case efficiency loss when compared to all other feasible mechanisms fitting the assumptions. To this end, we first define the class, $\mathcal{M}$, of mechanisms that we want to consider.

Definition 3: The class $\mathcal{M}$ of mechanisms consists of all supply functions, $M(b, p)$, such that the following conditions are satisfied:

1) *M defines a smooth market-clearing mechanism. Here, a differentiable $M : (0, \infty) \times \mathrm{R}^+ \to \mathrm{R}^+$ is said to be a smooth market clearing mechanism if for all $d > 0$, for all $n = |N| > 1$, and for all non-zero $b = (b_1, \ldots, b_N)$, $\exists$ a unique solution $p > 0$ to*

$$\sum_{i}^{n} M(b_i, p) = d. \tag{10}$$

2) *For all $C_i \in \mathcal{C}$, for all $u \in \mathcal{U}$, and for all $d > 0$, a DH's payoff is concave if it is benefit anticipating. $\mathcal{C}$ is the set consisting of all continuous, convex, and strictly increasing cost functions.*
3) *For all $C_i \in \mathcal{C}$, for all $u \in \mathcal{U}$, and for all $d > 0$, there exists a $b \geq 0$ such that $M(b_i, p) = q_i(b_i, p)$, $\forall i$.*

The second condition allows us to characterize Nash equilibria in terms of only the first-order conditions. To justify this condition, we note that some assumption of quasiconcavity is generally used to guarantee the existence of pure-strategy Nash equilibria [50]. The third condition ensures that given a benefit p and given $q_i(b_i, p) \in [0, d]$, each DH i can make a choice b_i to guarantee $q_i(b_i, p)$ - ensuring all possible demands can be chosen any market-clearing benefit. *In view of these conditions, it is evident that the class of mechanisms in $\mathcal{M}$ fit the privacy trading scenario we address in this work.* In this regard, we showcase the optimality of our proposed parametric mechanism, an element of the set $\mathcal{M}$, via the following theorem, the proof of which is in the Section VIII.

Theorem 6: Given $M \in \mathcal{M}$, the following results hold:

1) *There exists a competitive equilibrium b for any privacy trading market characterized by the triplet (d, N, U), where d is the total privacy compromise demand on the ad-network side, N is the number of competing DHs, and U is the vector of utility functions for every DH. Moreover, for any such b, the resulting privacy compromises, $q_i(b_i, p)$, for each DH i maximizes welfare.*
2) *There exists $B : (0, \infty) \to (0, \infty)$, a concave, strictly increasing, differentiable, and invertible function, such that for all $p > 0$, and $b_i \geq 0, \forall i \in N$, we have $M(b_i, p) = b_i B(p)$.*
3) *The worst case market efficiency loss under oligopoly is minimized if $M(b_i, p) = \Delta b_i p$, for some $\Delta > 0$.*

Theorem Implication - For privacy trading oligopoly markets, the linear supply function mechanism minimizes the loss in worst case market efficiency.

VI. COMPUTATIONAL EVALUATION

In this section, we focus on developing supply function bidding algorithms that converge in practice to market equilibria for perfectly competitive and oligopolistic markets in a distributed fashion. Our primary performance metric is market equilibrium convergence speed in terms of the number of iterations. Our motivation for coming up with distributed algorithms is the fact that DH cost functions are private information not released to an ad-network, and as a result the latter cannot centrally solve the optimization problems to maximize utilitarian social welfare and arrive at ONE, respectively. In addition, we need algorithms that are light on computation and communication overhead.

A. MINI REAL-WORLD EVALUATION SETUP

As part of a mini-experiment to evaluate supply function bidding algorithms, we collect sanitized consumer data for 1000 clients on their two sleep patterns (i.e., time to go to sleep, hours of sleep) from three fitness app startup firms A, B, and C based in northern California, USA. We ensure that the set of 1000 clients for each company do not overlap. For the aggregate data collected from both the companies, we set up an independent (of A, B, and C) sleep expert representative from a medical department at an university in northern California to *act* as an ad-network. The expert has thirty 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 in the fitness and pharmaceutical industries. Having collected real-world data, as a mock experiment, we synthetically implement a triopoly competition between A, B, and C by choosing a senior representative from both the firms to trade on the sanitized data of their clients with the ad-network, i.e., the medical representative, in return for (a) fictitious (but scaled on medical value of the data) monetary benefits and (b) some health insights on the available consumer data to be passed on by the representatives of A, B, and C to their clients. We emphasize here that the ad-network does not have knowledge of individual consumers whose data is under trade. Trading is done using the supply function mechanism and each of A, B, and C choose parameters of 1, 1, and 2 respectively, with a common demand upper limit of 100 differential privacy (DP) units, and a zero lower limit. Each DP unit is assumed to be 0.02. Each DH reports a nearly linear cost function to be of the form $C_i(q_i) = a_i q_i + h_i q_i^2$ with $a_i \geq 0$ and $h_i \ll a_i \geq 0$. More specifically, a_i values chosen by firms A, B, and C are 0.1, 0.2, and 0.1 respectively. Correspondingly, the h_i values chosen are 0.002, 0.005, and 0.005 respectively.

B. DISTRIBUTED BIDDING ALGORITHMS

As potential distributed algorithm candidate types, one could either use the standard dual gradient algorithm proposed in [51], or the alternative direction multiplier method in [52]. Both types are iterative in nature, and equivalently maps the supply bidding process. In this work[15] we resort to the dual gradient algorithm in [51], without loss of generality.

[15]We do not focus on the design of optimal distributed algorithms in terms of speed and scalability. Our goal is to just show fast convergence and scalability promise of implemented markets induced by supply function theory, and our proposed algorithms achieve them using as basis, the seminal algorithm type in [51].[16]

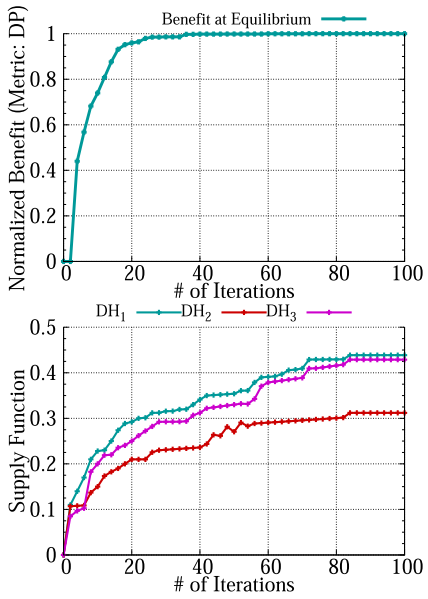


FIGURE 3. **(Benefit, Supply Function) at Market Equilibrium.**

The basic idea behind the two algorithms (see Algorithms 1 and 2 for perfectly competitive and oligopolistic markets, respectively) is the iterative interplay (until convergence) between the ad-network announcing a benefit p to the DHs, and the DHs subsequently updating their non-private bidding functions b_i to the ad-network. (see Figure 4 for a flowchart representation) In principle, the crux lies behind convergence lies in the Lagrangian of Equation (7) being strictly concave and thereby using the Projection Theorem [55] we arrive at the optimal benefit and supply functions at market equilibrium. Consequently, our proposed distributed bidding algorithms possess all the convergence properties of dual gradient algorithms. We refer the readers to [51] for details regarding the theory of optimal step sizes, the stopping criterion, and convergence speed. As an example of the high convergence speed, we show via experiments in the following section that for very low γ values in Algorithms 1 and 2, convergence is very fast, thereby showing great potential to ensure the property of scalability for large number of DHs. To be more specific, it is shown in [51] that in theory very small γ values result in an exponential convergence rate.

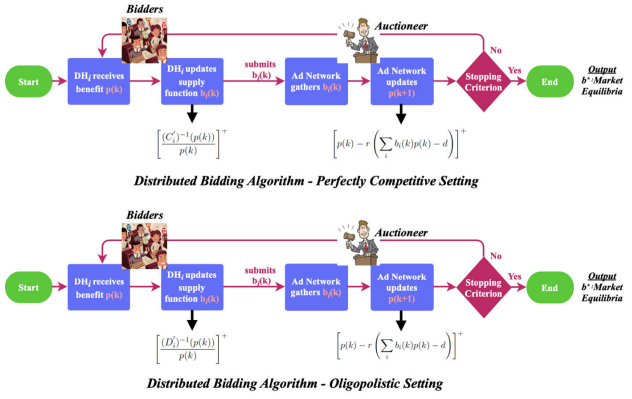


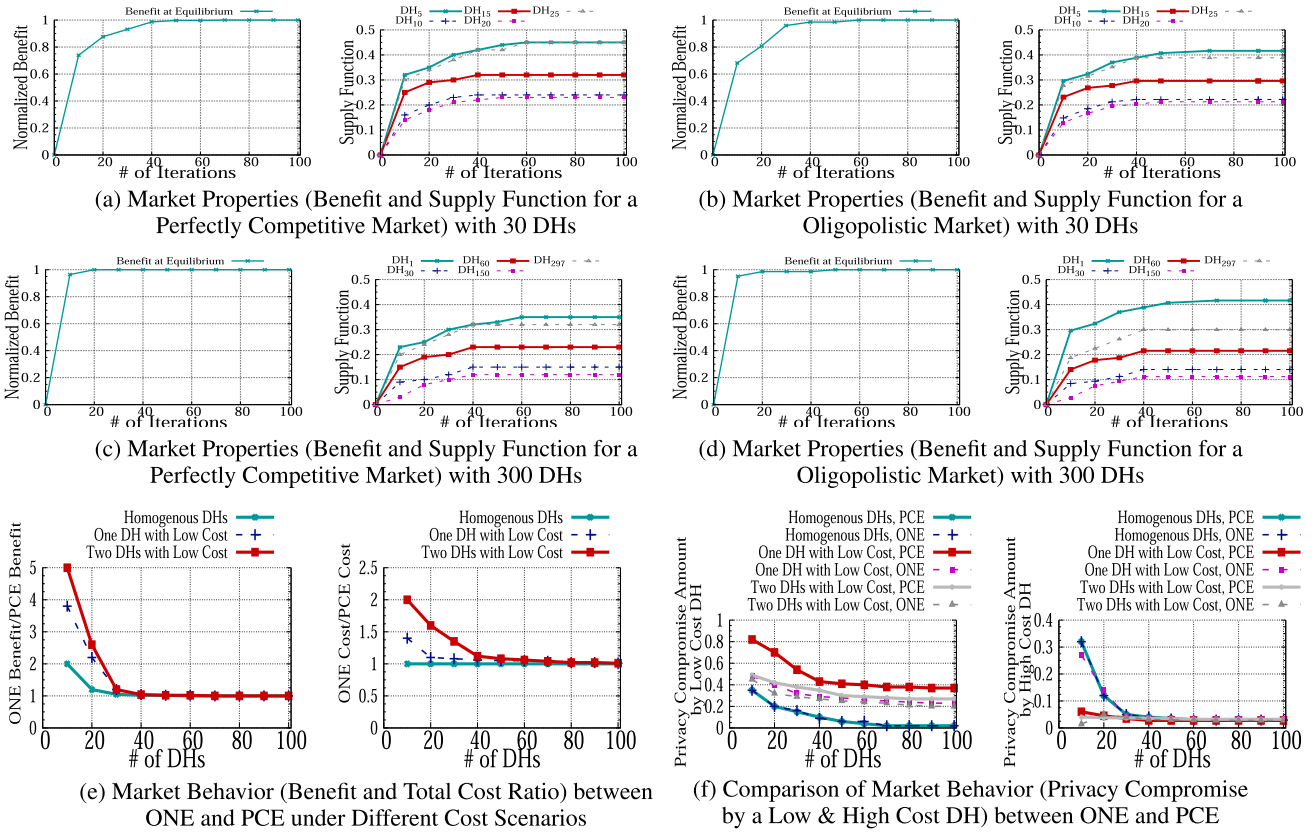
FIGURE 4. **Flowchart of Distributed Market Bidding Algorithms.**

C. SYNTHETIC EVALUATION SETUP

Due to a lack of real-world data, We experiment on *synthetic data*, to further investigate the scalability, efficiency, and convergence properties of our proposed algorithmic market mechanisms. We consider two DH population settings for our evaluations: (i) a privacy compromise setting with 30 DHs, and (ii) a significantly larger population setting with 300 DHs. For each DH i, we consider its cost function to be of the form $C_i(q_i) = a_i q_i + h_i q_i^2$ with $a_i \geq 0$ and $h_i \geq 0$. The reason for choosing cost functions of such types is their widespread use and popularity in economics (and also somewhat evident from our experiment with the app firms) due to (a) marginal costs can become either constant (when $h_i = 0$) or linear (when $h_i > 0$) with the amount of commodity in question, i.e., in our case the amount of privacy compromise, and this is reflective of practical microeconomic commodity settings (b) provides a very good approximation to higher order cost functions, if they were to exist. As a representative example (without loss of generality), for the 30 DH and 300 DH case respectively, the value of d is chosen to be 15 units (indicative of a low aggregate compromise) and 150 units (indicative of a high aggregate compromise) of a normalized information-theoretic privacy leakage metric[17] [10] we define to be $\frac{MI(X_i;Y_i)}{H(X_i)}$, where X_i is the source distribution[18] at the DH i and Y_i is the distribution at the ad-network of X_i, and $H(X_i)$ is the Shannon (information-theoretic) entropy of X_i, and $MI(X_i; Y_i)$ is the mutual information between X_i and Y_i. Note that $0 \leq \frac{MI(X_i;Y_i)}{H(X_i)} \leq 1$. a_i and h_i are randomly drawn without loss of generality from [1, 2] and [0, 4.5] respectively. We emphasize here that the constants chosen for our work is with the mindset that we can have DH cost functions taking low values and otherwise. Scaling up or down the constant range would not affect results as long as we have cost functions taking required value ranges. To study the impact of the DH cost functions on the efficiency loss in the ONE, we consider three cases: (i) DHs are homogeneous (a_i and h_i equals 1 and 2 respectively for all DH i), (ii) one DH has an extremely low cost function, and the other DHs have the same cost function, and (iii) two DHs have extremely low cost functions, and the other DHs have the same cost functions. For the low cost cases, we assume coefficients a_i and h_i to be 0.1 and 0.2 respectively for low cost DHs, while

[17] Our methodology is general and independent of the information-theoretic privacy metric. Differential Privacy is an example of an information-theoretic metric. In addition, one of the reasons for experimenting with a privacy metric different from that of differential privacy as in the real world case, is to test for consistence of results with different information-theoretic metrics.

[18] Consumer information collected by DHs can be represented as discrete or continuous random variables.



(a) Market Properties (Benefit and Supply Function for a Perfectly Competitive Market) with 30 DHs

(b) Market Properties (Benefit and Supply Function for a Oligopolistic Market) with 30 DHs

(c) Market Properties (Benefit and Supply Function for a Perfectly Competitive Market) with 300 DHs

(d) Market Properties (Benefit and Supply Function for a Oligopolistic Market) with 300 DHs

(e) Market Behavior (Benefit and Total Cost Ratio) between ONE and PCE under Different Cost Scenarios

(f) Comparison of Market Behavior (Privacy Compromise by a Low & High Cost DH) between ONE and PCE

FIGURE 5. Comparison of Market Properties with Linear Marginal DH Cost (Based on Synthetic Data).

others have their a_i and h_i coefficients set high and randomly selected in the interval [1, 2].

D. EVALUATION RESULTS

For our real-world experimental setting, we show in Figure 3 the results for benefit and supply function values at market equilibrium with respect to the number of iterations to market convergence. We observe that benefit and supply functions converge fast (within 25 iterations on a latest MacBook Pro with 16GB RAM) to the market equilibrium (ONE). This indicates the possibility of the existence of working markets satisfying all concerned stakeholders (as per our model) if personal data were to be traded. *As part of future plans, we would like to run larger scale field experiments, conditioned on the availability of real data, to validate our speed and scalability claims on working privacy trading markets.* However, in the absence of real-world data, we experiment with synthetic data as curated in Section VI.C. *Without loss of generality (and in the interest of space), we represent one of the 50 random instances in our plots. We not later on the rationale of not showing confidence interval bars in the plots.*

For purely synthetic settings, we observe from Figure 5b (where $\gamma = 0.1$, and DH marginal costs are linear) that benefit and supply functions in the 30 DH case converge fast (within 60 iterations on a latest MacBook Pro with 16GB RAM) to the market equilibrium (ONE). In addition, the benefit at ONE is higher than that in PCE - consistent with Theorem 5. Compared to the b_i value at PCE, DHs with low bids at the PCE tend to bid higher at the ONE, whereas DHs who have high bids at the PCE tend to bid a low value at ONE. The rationale here is that if a DH bids a low value at PCE, it has an incentive to bid higher at ONE because the benefit at ONE is higher and the DH might gain more. On the contrary, if a DH bids high at PCE, it may have an incentive to decrease bid at PCE because it might gain more by reducing privacy compromise amount but collecting the same benefit due to higher benefit at ONE. Through Figures 5c and 5d (where $\gamma = 0.05$), we show the scalability of Algorithms 1 and 2. *The results and rationale are similar to those in Figure 5b, and convergence to market equilibrium is equally fast.*

Figure 5e plots the comparison of benefit and total cost respectively at PCE and ONE. Figure 5f plots the amount of privacy compromise by low and high compromise cost users (denoting trade of low and high privacy sensitive data) respectively, at PCE and ONE. We observe form Figure 5e that if all DHs are homogeneous (i.e., trade data of similar privacy sensitivity), the differences between the market equilibrium benefits are small and the utilitarian social welfare of the two market equilibria are the same - consistent with Corollary 3. In all the three cases related to studying DH cost impact on efficiency loss mentioned in Section VI.C,

Algorithm 1 Distributed Bidding Algorithm - Perfectly Competitive Setting

1: On receiving benefit $p(k)$ announced by the ad-network, each DH_i updates its supply function, $b_i(k)$ according to

$$b_i(k) = \left[\frac{(C_i')^{-1}(p(k))}{p(k)}\right]^+ \tag{11}$$

and submits it to the ad-network. Here "+" denotes the projection onto $\mathbf{R}^+$, the set of non-negative real numbers.

2: On gathering bids $b_i(k)$ from DHs, the ad-network updates the benefit according to

$$p(k+1) = \left[p(k) - r\left(\sum_i b_i(k)p(k) - d\right)\right]^+ \tag{12}$$

and announces the benefit $p(k+1)$ to the DHs, where $r > 0$ is a constant stepsize.

3: Set $k \to k+1$

4: Check stopping criterion as mentioned in [51], and repeat

we observe from Figure 5f that the differences in efficiency loss between market equilibria decrease quickly with increase in the number of DHs. This is due to the fact that with increase in market size (multiple traders selling similar data), the market power of each DH decreases and oligopoly tends towards behaving like a perfectly competitive market. When the market size is small, the differences in loss between the two market equilibria (PCE and ONE) are large when only one DH has a low cost function - this is because the latter has market power (to attract customers who care for privacy). However, when two DHs have low cost functions the difference between the two market equilibria decreases rapidly, implying the fact that the ad-network or a regulator needs to introduce trading tiers of similar cost competing DHs in the market to improve social welfare. When the market size is large, the differences between the two market equilibria are small for all the three efficiency loss study cases cited in Section VI.C. However, as an interesting observation, for the case when two DHs have low cost functions, the benefit and cost ratio between two market equilibria is larger than in the case when only one DH has a low cost function. This is because all high cost DHs together contribute to a large fraction of the total privacy compromise amount, which limits the market power of the low cost DH. Thus, given a fixed large market size, low cost DHs in the two low-cost DH case will have a larger market power than the low cost DH in a single low-cost DH case, leading to a larger benefit and cost ratio. DHs facing low trading cost compromise less on privacy at ONE than in PCE, whereas DHs with high cost compromise more at ONE than in PCE. This is because at ONE, DHs have market power to increase the benefit. Low cost DHs gain more net revenue by decreasing their compromise amount, whereas high cost DHs have an incentive to compromise more privacy due to increased benefit.

The results for the case when *DH marginal costs are constant* is very similar and is shown through Figure 6b-6f. For such plots the a_i values are kept the same as in the case of linear marginal DH costs, and the h_i values are equal to zero. The reasoning behind the figures is the same as for Figure 5b-5f. *It is important to note that due to similarity of results for the instances, and the convergence (as visible through the plots), we do not need to show confidence intervals for the plots.*

Algorithm 2 Distributed Bidding Algorithm - Oligopolistic Setting

1: On receiving benefit $p(k)$ announced by the ad-network, each DH_i updates its supply function, $b_i(k)$ according to

$$b_i(k) = \left[\frac{(D_i')^{-1}(p(k))}{p(k)}\right]^+ \tag{13}$$

and submits it to the ad-network. Here "+" denotes the projection onto $\mathbf{R}^+$, the set of non-negative real numbers.

2: On gathering bids $b_i(k)$ from DHs, the ad-network updates the benefit according to

$$p(k+1) = \left[p(k) - r\left(\sum_i b_i(k)p(k) - d\right)\right]^+ \tag{14}$$

and announces the benefit $p(k+1)$ to the DHs, where $r > 0$ is a constant stepsize.

3: Set $k \to k+1$

4: Check stopping criterion as mentioned in [51], and repeat

VII. SUMMARY AND FUTURE WORK

In this paper, we proposed a introductory but rigorous preference-based privacy trading market model for mobile in-app ecosystems of the current data surveillance age that aims to achieve a maximum privacy welfare state amongst competing data holders (e.g., apps) by preserving their *heterogeneous* privacy preservation constraints upto certain compromise levels (in return for benefits to data holders), induced by their clients, and at the same time satisfying requirements of agencies (e.g., advertisers) that collect client data for the purpose of targeted advertising. *More importantly, our proposed trading methodology is consensual in the sense that pre-trading, DHs can decide on their trading preferences as a function of the benefit to be offered, without needing to sell non-voluntarily with no explicitly offered benefit.* To this end, using concepts from supply-function economics, we proposed the first mathematically rigorous privacy market design paradigm with private DH cost functions that characterized states of market efficiency as well as inefficiency by respecting *heterogeneous privacy constraints* of competing data holders to extents possible, in a provably optimal fashion. More specifically, we analyzed perfectly competitive and oligopolistic markets to achieve market equilibria that is efficient in the former, but not in the latter

due to negative externalities of trading not being internalized. Consequently, we characterized the efficiency gap in closed form. As a major finding, we showed that increasing competition between app firms of similar market power for privacy trading activities contribute to increased economic social welfare due to trading externalities being internalized better between similar firm types, thereby suggesting regulators to enable privacy trading in segregated pools of similar app firms.

As part of future work, we plan to (a) gauge the preference supply functions of individual DHs using large-scale social experiments, and (b) investigate the existence of efficient/boundedly inefficient multi-supplier (apps), multi-demand side (ad-exchanges) market competition models in a privacy trade setting, and explicitly account for information correlations between supplier side data.

VIII. PROOFS OF THEOREMS

Proof of Theorem 1: Definition 1 tells that $\{(\bar{b}_i)_{i\in N}, \bar{p}\}$ is a competitive equilibrium if and only if

$$(C_i'(q_i(\bar{b}_i, \bar{p}_i)) - \bar{p})(b_i - \bar{b}_i) \geq 0, \ \forall b_i \geq 0 \tag{15a}$$

$$\sum_i q_i(\bar{b}_i, \bar{p}) = d \tag{15b}$$

Here, (15a) results from the optimality condition of the convex optimization problem of DH net revenue, and (15b) follows directly from Definition 1. Since $\bar{p} \geq 0$, multiplying $\bar{p}$ to (15a), we get

$$(C_i'(\bar{q}_i) - \bar{p})(q_i - \bar{q}_i) \geq 0, \quad \forall q_i \geq 0 \tag{16a}$$

$$\sum_i \bar{q}_i = d \tag{16b}$$

This is just the KKT optimality condition of the optimization problem in the theorem. Hence, $(q_i)_{i\in N}$ maximizes social welfare. And if $\{(\bar{q}_i)_{i\in N}, \bar{p}\}$ is an optimal solution of the latter optimization problem, $\{\left(\bar{b}_i = \frac{\bar{q}_i}{\bar{p}}\right)_{i\in N}, \bar{p}\}$ satisfies (15a); this tells that $\{(\bar{b}_i)_{i\in N}, \bar{p}\}$ is a competitive equilibrium. If $C_i(q_i)$ is convex for each DH i, then the social welfare maximization problem is a strictly convex problem. Thus there exists a unique optimal solution $(\bar{q}_i)_{i\in N}$. Moreover, from (16a), $\bar{p} = C_i'(\bar{q}_i)$ for any $\bar{q}_i \geq 0 \Rightarrow \bar{p}$ is unique $\Rightarrow$ unique equilibrium. ■

Proof of Theorem 2: From the proof of Theorem 1, we know that $\{\bar{p}, (\bar{q}_i)_{i\in N}\}$ satisfies (16a) and (16b). From (16a), we know that, for any $i \in N$, 1) if $\bar{q}_i > 0$, then $\bar{p} = C_i'(\bar{q}_i) \geq C_i'(0)$, 2) if $\bar{q}_i = 0$, then $\bar{p} \leq C_i'(\bar{q}_i) = C_i'(0)$. Thus, we know all the DHs who compromise on privacy have a smaller $C_i^* = C_i'(0)$ than those who do not. Since C_i^* is increasing in i, $\bar{N}$ takes the form of $1, 2, \ldots, \bar{n}$. If $\bar{n} < |N|$, then 1 and 2 imply that $C_{\bar{n}}^0 \leq \bar{p} \leq C_{\bar{n}+1}^0$. If $\bar{n} = |N|$, $\bar{p} = C_{|N|}'(\bar{q}_{|N|}) \leq C_{|N|}'(d) = C_{n+1}^0$, thus $C_{\bar{n}}^0 \leq \bar{p} \leq C_{n+1}^0$. Note that, $C_i'(q_i')$ is an increasing function. Hence $\sum_i^{\bar{n}}(C_i')^{-1}(C_{\bar{n}}^0) \leq \sum_i^{\bar{n}}(C_i')^{-1}(\bar{p}) \leq \sum_i^{\bar{n}}(C_i')^{-1}(C_{\bar{n}+1}^0)$ which is $\sum_i^{\bar{n}}(C_i')^{-1}(C_{\bar{n}}^0) \leq \sum_i^{\bar{n}} \bar{q}_i = d \leq \sum_i^{\bar{n}}(C_i')^{-1}(C_{\bar{n}+1}^0)$. ■

Proof of Corollary 1: From Theorem 2, we know that $\forall i \in \bar{N}, \bar{p} = C_i'(\bar{q}_i)$. Notice that $C_i(\cdot)$ is a convex function. Thus $C_i(\bar{q}_i) - C_i(0) \leq C_i'(\bar{q}_i)\bar{q}_i$. As $C_i(0) = 0$, we have $C_i(\bar{q}_i) \leq \bar{p}\bar{q}_i$. ■

Proof of Lemma 1: We prove the result by contradiction. Suppose that it does not hold, and without loss of generality, assume that $\sum_{j\neq 1} b_j^* = 0$ for DH_i. Then the payoff for the DH_i is $U_i(b_i^*, b_{-i}^*) = 0$ if $b_i^* = 0$, and $U_i(b_i^*, b_{-i}^*) = \frac{d^2}{b_i^*} - C_i(d)$ if $b_i^* > 0$. We see that when $b_i^* = 0$, DH_i has an incentive to increase it, and when $b_i^* \geq 0$, DH_i has an incentive to decrease it. So, there is no Nash equilibrium with $\sum_{j\neq i} b_j^* = 0$. ■

Proof of Lemma 3: We have

$$\begin{aligned} U_i(b_i, b_{-i}) &= p(b)q_i(p(b), b_i) - C_i(q_i(p(b), b_i)) \\ &= \frac{d^2 b_i}{\left(\sum_j b_j\right)^2} - C_i\left(\frac{db_i}{\sum_j b_j}\right) \end{aligned} \tag{17}$$

From (17), we have

$$\begin{aligned} &\frac{\partial U_i(b_i, b_{-i})}{\partial b_i} \\ &= \frac{d^2\,(B_{-i} - b_i)}{(B_{-i} + b_i)^3} - \frac{dB_{-i}}{(B_{-i} + b_i)^2} C_i'\left(\frac{db_i}{B_{-i} + b_i}\right) \\ &= \frac{d^2}{(B_{-i} + b_i)^2}\left[\frac{B_{-i} - b_i}{B_{-i} + b_i} - \frac{B_{-i}}{d} C_i'\left(\frac{db_i}{B_{-i} + b_i}\right)\right] \end{aligned} \tag{18}$$

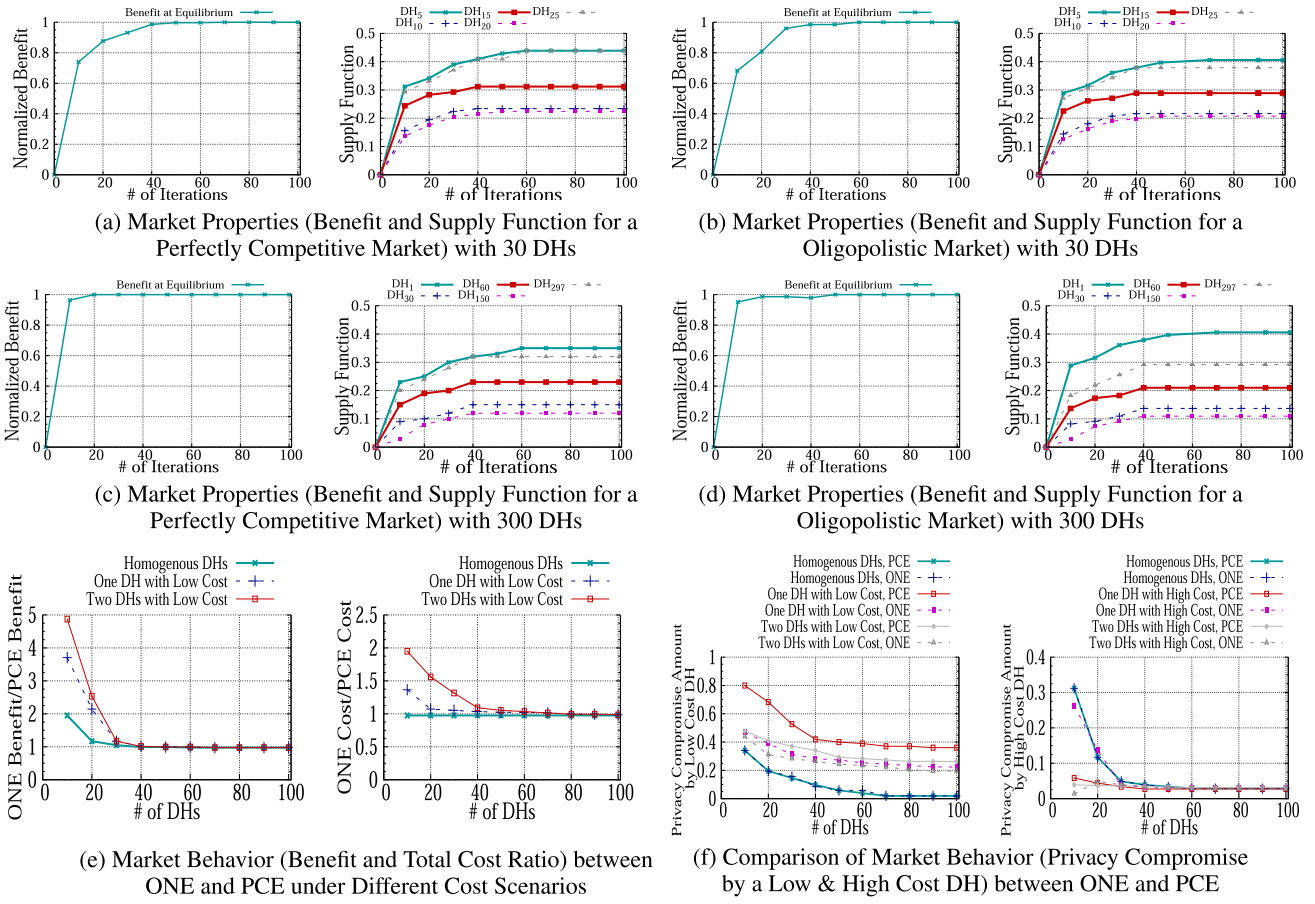
The first form in the square bracket in (18) is no greater than 1 and strictly decreasing in b_i, the second term is increasing in b_i. So, if $\frac{B_{-i}}{dC_i'(0)} \geq 1$ and $\frac{\partial U_i(b_i, b_{-i})}{\partial b_i} \leq 0 \ \forall b_i$, and $b_i = 0$ maximizes $DH_i's$ payoff $U_i(b_i, b_{-i})$ for the given b_{-i}. If $\frac{B_{-i}}{dC_i'(0)} \leq 1, \frac{\partial U_i(b_i, b_{-i})}{\partial b_i} = 0$ only at one point $b_i > 0$. Furthermore, note that $\frac{\partial U_i(0, b_{-i})}{\partial b_i} > 0$ and $\frac{\partial U_i(B_{-i}, b_{-i})}{\partial b_i} \leq 0$. So, the point b_i maximizes $DH_i's$ payoff $U_i(b_i, b_{-i})$ for a given b_{-i}. Thus, at Nash equilibrium, b^*,

b^*satisfies

$$\begin{cases} b_i^* = 0, \forall i, \text{ if } \dfrac{B_{-i}^*}{dC_i'(0)} \geq 1 \\ \dfrac{B_{-i}^* - b_i^*}{B_{-i}^* + b_i^*} - \dfrac{B_{-i}^*}{d} C_i'\left(\dfrac{db_i^*}{B_{-i}^* + b_i^*}\right) = 0, \text{ otherwise} \end{cases} \tag{19}$$

Given a Nash equilibrium, b^*: 1) if $b_i^* = 0$, then $b_i^* < B_{-i}^*$ from lemma 1 and, 2) otherwise, b_i^* satisfies (19). Note that the second term on the left hand side of (19) is positive. So the first term must be positive as well, which requires $B_{-i}^* > b_i^*$. Because for each DH_i, $q_i^* = \frac{b_i^* d}{b_i^*} + B_{-i}^*$, each DH will compromise a privacy of less than $\frac{d}{2}$ at the equilibrium. ■

Proof of Theorem 3: Here, we prove the existence and uniqueness of the optimal solution of optimization problem in Theorem 3. We first pick $\hat{d} < \frac{d}{2}$ such that $|N| \cdot \hat{d} > d$ and solve this problem: $\min_{0\leq q_i<\hat{d}} \sum_i D_i(q_i)$ subject to



(a) Market Properties (Benefit and Supply Function for a Perfectly Competitive Market) with 30 DHs

(b) Market Properties (Benefit and Supply Function for a Oligopolistic Market) with 30 DHs

(c) Market Properties (Benefit and Supply Function for a Perfectly Competitive Market) with 300 DHs

(d) Market Properties (Benefit and Supply Function for a Oligopolistic Market) with 300 DHs

(e) Market Behavior (Benefit and Total Cost Ratio) between ONE and PCE under Different Cost Scenarios

(f) Comparison of Market Behavior (Privacy Compromise by a Low & High Cost DH) between ONE and PCE

FIGURE 6. **Comparison of Market Properties with Constant Marginal DH Cost (Based on Synthetic Data).**

$\sum_i q_i = d$. Denote optimal value of this problem as D_d^*. For each i, find ε_i such that $D_i(q_i) \geq D_d^*$ for all $q_i \in \left[\frac{d}{2} - \varepsilon_i, \frac{d}{2}\right)$. Such ε_i always exists because $D_i(q_i)$ is a strictly increasing function and $\lim_{q_i \to \frac{d}{2}} D_i(q_i) = \infty$. Therefore, we confer that the optimization problem in Theorem 3 is equivalent to this problem: $\min_{0 \leq q_i \leq \frac{d}{2} - \varepsilon_i} \sum_i D_i(q_i)$ subject to $\sum_i q_i = d$, which has a unique solution. Therefore, the optimal solution always exists and the uniqueness follows from strict convexity of $D_i(q_i)$.

Now we first note that

$$D_i'(q_i) = \left(1 + \frac{q_i}{d - 2q_i}\right) C_i'(q_i) \tag{20}$$

which is positive, strictly increasing function in $q_i \in \left[0, \frac{d}{2}\right)$. So, $D_i(q_i)$ is strictly increasing and strictly convex function in $\left[0, \frac{d}{2}\right)$ because $D_i(q_i) = \int_0^{q_i} D_i'(x_i)dx_i \geq C_i'(0) \int_0^{q_i} \left(1 + \frac{x_i}{d} - 2x_i\right) dx_i = C_i'(0) \int_0^{q_i} (\frac{1}{2} + \frac{d}{2d} - 2x_i)dx_i = C_i'(0) \int_0^{q_i} \left(\frac{1}{2q_i} - \frac{d}{4log(d-2x_i)}\right) dx_i$. Thus, $\lim_{q_i \to \frac{d}{2}} D_i(q_i) = \infty$. Therefore, the optimization problem in the theorem is strictly convex problem and has unique optimal solution, and after a bit of mathematical manipulation, we get the unique solution q^* determined by

$$\left(p^* - \left(1 + \frac{q_i^*}{d - 2q_i^*}\right) C_i'(q_i^*)\right)(q_i - q_i^*) \leq 0, \forall q_i \tag{21a}$$

$$\sum_i q_i^* = d \tag{21b}$$

$$p^* > 0 \tag{21c}$$

$$\left(\frac{d}{B_{-i}^* + b_i^*} - \frac{B_{-i}^*}{B_{-i}^* - b_i^*} C_i'\left(\frac{db_i^*}{B_{-i}^* + b_i^*}\right)\right)(b_i - b_i^*) \leq 0, \forall b_i \tag{21d}$$

Recall that the the Nash equilibrium value of $p^* = \frac{d}{\sum_i b_i^*}$ and the corresponding Nash equilibrium allocation $q_i^* = b_i^* p^*$. We can write (21d) as $\left(p^* - \left(\frac{q_i^*}{d-2q_i^*}\right) C_i^*(q_i^*)\right)(b_i p^* - q_i^*) \leq 0$. Note that at the Nash equilibrium, $p^* > 0$ since $\sum_i b_i^* > 0$ by lemma 1. Thus the Nash equilibrium of the game satisfies (21a) - (21c), and solves the optimization problem in the theorem. The existence and uniqueness of the Nash equilibrium is a result of the existence and uniqueness of the optimal solution of the optimization problem. ■

Proof of Theorem 4: Note that $D_i'(q_i)$ is a strictly increasing function of q_i and $D_i'(0) = C_i'(0)$. The proof follows the same argument as in Theorem 2. ■

Proof of Corollary 3: From Theorem 4, we know that $\forall i \in \bar{N}, p^* = D_i'(q_i^*)$. Notice that $D_i(\cdot)$ is a strictly convex function. Thus, $D_i(q_i^*) - D_i(0) < D_i'(q_i^*)q_i^*$. Because $D_i(0) = 0$, $D_i(q) > C_i(q)$, we have $C_i(q_i^*) < p^*q_i^*$. ■

Proof of Theorem 5: Notice that $D_i'(q_i)$ and $C_i'(q_i)$ are both strictly increasing function and $D_i'(q_i) \geq C_i'(q_i)$ for any $q_i \in \left[0, \frac{d}{2}\right)$. For any $i \in N$, $(D_i')^{-1}(\bar{p}) \leq C_i^{-1}(\bar{p})$. Suppose $p^* < \bar{p}$. Because $C_{n^*}^0 \leq p^* \leq C_{n^*+1}^0$, $C_{\bar{n}}^0 \leq \bar{p} \leq C_{\bar{n}+1}^0$, and $C_1^0 \leq C_2^0 \leq \ldots\ldots \leq C_n^0$, we have $n^* \leq \bar{n}$. Therefore, $\sum_i^{n^*}(D_i')^{-1}(p^*) < \sum_i^{n^*}(D_i')^{-1}(\bar{p}) \leq \sum_i^{n^*}(C_i')^{-1}(\bar{p}) \leq \sum_i^{\bar{n}}(C_i')^{-1}(\bar{p}) = d$, which contradicts that $\sum_i^{n^*}(D_i')^{-1}(p^*) = d$. Thus, $p^* \leq \bar{p}$. Therefore, $\bar{n} \leq n^*$, implying $\bar{N} \subset N^*$. If $n^* < n$, then $p^* \leq D_{n^*+1}'(0) \leq D_{n^*+1}'\left(\frac{d}{n}\right) = \frac{n-1}{n-2}C_{n^*+1}'\left(\frac{d}{n}\right) \leq \frac{n-1}{n} - 2M$. If $n^* = n$, there exists one DH_j such that $0 < q_j^* \leq \frac{d}{n}$. Thus, $p^* = D_j'(q_i^*) \leq D_j\left(\frac{d}{n}\right) \leq \frac{n-1}{n} - 2M$. In summary,

$$p^* \leq \frac{n-1}{n-2}M \tag{22}$$

On the other side, there exists at least one DH_j such that $C_j'(\bar{q}_i) = \bar{p}$ and $\bar{q}_i \geq \frac{d}{n}$. Thus,

$$\bar{p} \geq C_j'\left(\frac{d}{n}\right) \geq m \tag{23}$$

Combing (22) and (23) gives $p^* \leq \frac{n-1}{n-2}\frac{M}{m}\bar{p}$. Lastly, $\bar{C} \leq C^*$ comes from the fact that $(\bar{q}_i)_{i\in N}$ is an optimal solution of optimization problem in Theorem 1. If $\bar{q}_{max} < \frac{d}{2}$, then $\sum_i D_i(q_i^*) \leq \sum_i D_i(\bar{q}_i)$ since $(q^*)_{i\in N}$ is an optimal solution of optimization problem in Theorem 3. It is straightforward to check that

$$D_i(\bar{q}_i) \leq \left(1 + \frac{\bar{q}_i}{d} - 2\bar{q}_i\right) C_i(q_i^*).$$

Thus, $\sum_i D_i(q_i^*) \leq \left(1 + \frac{\bar{q}_{max}}{d-2\bar{q}_{max}}\right)\bar{C}$. On the other hand for any

$$\begin{aligned} q_i < \frac{d}{2}, \quad D_i(q_i) &= \left(1 + \frac{q_i}{d-2q_i}\right) C_i(q_i) \\ &\quad - \int_0^{q_i} \frac{d}{(d-2x_i)^2} C_i(x_i)dx_i \\ &\geq \left(1 + \frac{q_i}{d-2q_i}\right) C_i(q_i) - C_i(q_i)\int_0^{q_i} \frac{d}{(d-2x_i)^2}dx_i \\ &\geq \left(1 + \frac{q_i}{d-2q_i}\right) C_i(q_i) - C_i(q_i)\frac{q_i}{d-2q_i} \geq C_i(q_i). \end{aligned}$$

Thus,

$$C^* = \sum_i C_i(q_i^*) \leq \sum_i D_i(q_i^*) \leq \left(1 + \frac{\bar{q}_{max}}{d-2\bar{q}_{max}}\right)\bar{C}.$$

■

Proof of Theorem 6: The proof of this theorem is dealt in various steps, the first step recognizing that proof directly follows from Theorem 1 in [56] due to the similarity in structure.

Steps 2 of Proof of Theorem 6: A user's payoff is concave if he is price taking. The condition that a uniform market-clearing price must exist implies that for any fixed $\theta > 0$, the range of $D(\mu, \theta)$ must contain $(0, \infty)$ as μ varies in $(0, \infty)$. Now suppose that for fixed $\theta > 0$, there exist $\mu_1, \mu_2 > 0$ with $\mu_1 \neq \mu_2$ such that $D(\mu_1, \theta) = D(\mu_2, \theta) = d$, where $d > 0$. Let $C = 2d$ and let $R = 2$. Then for $\boldsymbol{\theta} = (\theta, \theta)$, there cannot exist a unique market-clearing price $p_D(\boldsymbol{\theta})$; so we conclude that $D(\cdot, \theta)$ is monotonic, and strictly monotonic in the region where it is nonzero.

Let $I \subset (0, \infty)$ be the set of $\theta > 0$ such that $D(\mu, 0)$ is monotonically nondecreasing in μ. From the preceding paragraph, we conclude that if $\theta \in (0, \infty)\backslash I$, then $D(\mu, \theta)$ is necessarily monotonically nonincreasing in μ. Further, if $\theta \in I$, then $D(\mu, \theta) \to \infty$ as $\mu \to \infty$, and $D(\mu, \theta) \to 0$ as $\mu \to 0$; on the other hand, if $\theta \in (0, \infty)\backslash I$, then $D(\mu, \theta) \to 0$ as $\mu \to \infty$, and $D(\mu, \theta) \to \infty$ as $\mu \to 0$.

Suppose $I \neq (0, \infty)$ and $I \neq \emptyset$; then choose $\theta \in \partial I$, the boundary of I. Choose a sequence $\theta_n \in I$ such that $\theta_n \to \theta$; and choose another sequence $\hat{\theta}_n \in (0, \infty)\backslash I$ such that $\hat{\theta}_n \to \theta$. Fix μ_1, μ_2 with $0 < \mu_1 < \mu_2$, such that $D(\mu_1, \theta) > 0$ and $D(\mu_2, \theta) > 0$. Then we have $D(\mu_1, \theta_n) \leq D(\mu_2, \theta_n)$, and $D(\mu_1, \hat{\theta}_n) \geq D(\mu_2, \hat{\theta}_n)$. Taking limits as $n \to \infty$, we get $D(\mu_1, \theta) \leq D(\mu_2, \theta)$, and $D(\mu_1, \theta) \geq D(\mu_2, \theta)$, so that $D(\mu_1, \theta) = D(\mu_2, \theta)$. But this is not possible, since $D(\cdot, \theta)$ must be strictly monotonic in the region where it is nonzero. Thus $I = (0, \infty)$ or $I = \emptyset$.

We will use Step 1 to show $D(\mu, \theta)$ is concave in $\theta \geq 0$ for fixed $\mu > 0$. Since $D(\mu, \theta)$ is continuous, it suffices to show that $D(\mu, \theta)$ is concave for $\theta > 0$. Suppose not; fix $\theta > 0, \overline{\theta} > 0$, and $\delta \in (0, 1)$ such that:

$$D(\mu, \delta\theta + (1-\delta)\overline{\theta}) < \delta D(\mu, \theta) + (1-\delta)D(\mu, \overline{\theta}) \tag{EC.1}$$

Note this implies in particular that either $D(\mu, \theta) > 0$ or $D(\mu, \overline{\theta}) > 0$. We assume without loss of generality that $D(\mu, \theta) > 0$. Let $C^R = RD(\mu, \theta)$, and let $\boldsymbol{\theta}^R = (\theta, \cdots, \theta) \in (\mathbb{R}^+)^R$. To emphasize the dependence of the market-clearing price on the capacity, we will let $p_D(\overline{\boldsymbol{\theta}}; C)$ denote the market-clearing price when the composite strategy vector is $\overline{\boldsymbol{\theta}}$ and the capacity is C. We will show that for any $\theta' > 0$, if $\mu^R = p_D(\boldsymbol{\theta}^{R-1}, \theta'; C^R)$, then $\mu^R \to \mu$ as $R \to \infty$. First note that by definition, we have $D(\mu^R, \theta') + (R-1)D(\mu^R, \theta) = RD(\mu, \theta)$; or, rewriting, we have:

$$\frac{1}{R}D(\mu^R, \theta') + \left(1 - \frac{1}{R}\right)D(\mu^R, \theta) = D(\mu, \theta) \tag{EC.2}$$

Now note that as $R \to \infty$, the right hand side remains constant. Suppose that $\mu^R \to \infty$. Since $I = (0, \infty)$ or $I = \emptyset$, either $D(\mu^R, \theta'), D(\mu^R, \theta) \to 0$, or $D(\mu^R, \theta'), D(\mu^R, \theta) \to \infty$; in either case, the equality (EC.2) is violated for large R. A similar conclusion holds if $\mu^R \to 0$ as $R \to \infty$. Thus we do not have $\mu^R \to 0$ or $\mu^R \to \infty$as $R \to \infty$.

Choose a convergent subsequence, such that $\mu_k^R \to \hat{\mu}$, where $\hat{\mu} \in (0, \infty)$. From (EC.2), we mush have $D(\hat{\mu}, \theta) = D(\mu, \theta)$. But as established above, since $D(\cdot, \theta)$ is strictly monotonic in the region where it is nonzero, this is only possible if $\hat{\mu} = \mu$. We conclude that the following three limits hold:

$$\lim_{R\to\infty} p_D(\boldsymbol{\theta}^R; C^R) = \mu;$$
$$\lim_{R\to\infty} p_D(\boldsymbol{\theta}^{R-1}, \overline{\theta}; C^R) = \mu;$$
$$\lim_{R\to\infty} p_D(\boldsymbol{\theta}^{R-1}, \delta\theta + (1-\delta)\overline{\theta}); C^R) = \mu; .$$

The remainder of the proof is straightforward. From (EC.1), for R sufficiently large, we must have:

$$D(p_D(\boldsymbol{\theta}^{R-1}, \delta\theta + (1-\delta)\overline{\theta}); C^R), \delta\theta + (1-\delta)\overline{\theta})$$
$$< \delta D(p_D(\boldsymbol{\theta}^R; C^R), \theta) + (1-\delta) D(p_D(\boldsymbol{\theta}^{R-1}, \overline{\theta}; C^R), \overline{\theta}).$$

This violates the conclusion of Step 1, so we conclude $D(\mu, \theta)$ is concave in $\theta \geq 0$ give $\mu > 0$. A similar argument shows that $\mu D(\mu, \theta)$ is convex in θ, by using the fact that $p_D(\boldsymbol{\theta}) D(p_D(\boldsymbol{\theta}), \theta_r)$ must be convex in θ_r for nonzero $\boldsymbol{\theta}$. Combining these results yields the desired conclusion.

Step 5, Proof of Theorem 6: B is an invertible, differentiable, strictly increasing, and concave function on $(0, \infty)$. Note from (10) that:

$$B(p_D(\boldsymbol{\theta})) = \frac{\sum_{r=1}^R \theta_r}{C}. \tag{EC.3}$$

We immediately see that B must be invertible on $(0, \infty)$; it is clearly onto, as the right hand side of (EC.3) can take any value in $(0, \infty)$. Furthermore, if $B(p_1) = B(p_2) = \gamma$ for some prices $p_1, p_2 > 0$, then choosing $\boldsymbol{\theta}$ such that $\sum_{r=1}^R \theta_r / C = \gamma$, we find that $p_D(\boldsymbol{\theta})$ is not uniquely defined. Thus B is one-to-one as well, and hence invertible. Finally, note that since D is differentiable, B must be differentiable as well. We let Φ denote the differentiable inverse of B. We will show that Φ is strictly increasing and convex. We first note that for nonzero $\boldsymbol{\theta}$ we have:

$$p_D(\boldsymbol{\theta}) = \Phi\left(\frac{\sum_{r=1}^R \theta_r}{C}\right).$$

Let

$$w_r(\boldsymbol{\theta}) = p_D(\boldsymbol{\theta}) D(p_D(\boldsymbol{\theta}), \theta_r)$$
$$= \Phi\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s} C\right) \tag{EC.4}$$

By Step 1, $w_r(\boldsymbol{\theta})$ is convex in $\theta_r > 0$. By considering strategy vectors $\boldsymbol{\theta}$ for which $\boldsymbol{\theta}_{-r} = 0$, it follows that Φ is convex.

It remains to be shown that Φ is strictly increasing. Since Φ is invertible, it must be monotonic; and thus Φ is either strictly increasing or strictly decreasing. To simplify the argument, we assume that Φ is twice differentiable. We twice differencetiate $w_r(\boldsymbol{\theta})$, given in (EC.4). Letting $\mu = \sum_{s=1}^R \theta_s / C$, we have for nonzero $\boldsymbol{\theta}$:

$$\frac{\partial^2 w_r}{\partial \theta_r^2}(\boldsymbol{\theta}) = \Phi''(\mu)\frac{\theta_r}{C^2\mu} + \frac{2\sum_{s\neq r}\theta_s}{C^2\mu^3}(\mu\Phi'(\mu) - \Phi(\mu)). \tag{EC.5}$$

Consider some nonzero $\boldsymbol{\theta}_{-r}$, and take the limit as $\theta_r \to 0$. The limit of the left-hand side in (EC.5) is nonnegative, by the convexity of $w_r(\boldsymbol{\theta})$ in $\theta_r > 0$. The limit of the first term in the right-hand side of (EC.5) is zero. Since $\Phi(\mu) > 0$, it follows that $\Phi'(\mu) > 0$, so that Φ is strictly increasing. This establishes the desired facts regarding B.

Steps 6, Proof of Theorem 6: Let (C,R,U) be a utility system. A vector $\boldsymbol{\theta} \geq 0$ *is a Nash equilibrium if and only if at least two components of* $\boldsymbol{\theta}$ *are nonzero, and there exists a nonzero vector* $\boldsymbol{d} \geq 0$ *and a scalar* $\mu > 0$ *such that* $\theta_r = \mu d_r$ *for all* r, $\sum_{r=1}^R d_r = C$, *and the following conditions hold:*

$$U_r'(d_r)\left(1 - \frac{d_r}{C}\right) = \Phi(\mu)\left(1 - \frac{d_r}{C}\right)$$
$$+\mu\Phi'(\mu)\left(\frac{d_r}{C}\right), \text{if } d_r > 0;$$
$$U_r'(0) \leq \Phi(\mu), \text{if } d_r = 0.$$

In this case $d_r = D(p_D(\boldsymbol{\theta}), \theta_r))$, $\mu = \sum_{r=1}^R \theta_r / C$, *and* $\Phi(\mu) = p_D(\boldsymbol{\theta})$. Suppose that $\boldsymbol{\theta}$ is a Nash equilibrium. Since $Q_r(\theta_r; \boldsymbol{\theta}_{-r}) = -\infty$ if $\boldsymbol{\theta} = 0$, (from (7)), we must have $\boldsymbol{\theta} \neq 0$. Suppose then that only one component of $\boldsymbol{\theta}$ is nonzero; say $\theta_r > 0$, and $\boldsymbol{\theta}_{-r} = 0$. Then the payoff to user r is:

$$U_r(C) - \Phi\left(\frac{\theta_r}{C}\right) C$$

But now observe that by infinitesimally reducing θ_r, user r can strictly improve his payoff (since Φ is strictly increasing). Thus $\boldsymbol{\theta}$ could not have been a Nash equilibrium; we conclude that at least two components of $\boldsymbol{\theta}$ are nonzero. In this case, from (7), and the expressions in (11) and (EC.4), the payoff $Q_r(\overline{\theta}_r; \boldsymbol{\theta}_{-r})$ to user r is differentiable. When two components of $\boldsymbol{\theta}$ are nonzero, we may write the payoff Q_r to user r as follows, using (11) and (EC.4):

$$Q_r(\theta_r; \boldsymbol{\theta}_{-r}) = U_r\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s} C\right)$$
$$-\Phi\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s} C\right).$$

Differentiating the previous expression with respect to θ_r, we conclude that if $\boldsymbol{\theta}$ is a Nash equilibrium then the following optimality conditions hold for each r:

$$F_r(\boldsymbol{\theta}) = 0 \text{if} \theta_r > 0; \tag{EC.6}$$
$$F_r(\boldsymbol{\theta}) \leq 0 \text{if} \theta_r = 0, \tag{24}$$

where

$$F_r(\boldsymbol{\theta}) = U_r'\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s} C\right)\left(\frac{C}{\sum_{s=1}^R \theta_s} - \frac{\theta_r C}{(\sum_{s=1}^R \theta_s)^2}\right)$$
$$-\Phi'\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{\theta_r}{\sum_{s=1}^R \theta_s}\right)$$
$$-\Phi\left(\frac{\sum_{s=1}^R \theta_s}{C}\right)\left(\frac{C}{\sum_{s=1}^R \theta_s} - \frac{\theta_r C}{(\sum_{s=1}^R \theta_s)^2}\right)$$

These conditions are equivalent to (14)-(15), if we make the substitutions $\mu = \sum_{s=1}^{R} \theta_s/C$, and $d_r = D(p_D(\boldsymbol{\theta}), \theta_r)$. Furthermore, in this case we have $\boldsymbol{d} \geq 0, \mu > 0, \theta_r = \mu d_r, \sum_{r=1}^{R} d_r = C$, and $p_D(\boldsymbol{\theta}) = \Phi(\mu)$.

On the other hand, suppose that we have found $\boldsymbol{\theta}, \boldsymbol{d}$ and μ such that the conditions of Step 6 are satisfied. In this case we simply reverse the argument above; since $Q_r(\overline{\theta}_r; \boldsymbol{\theta}_{-r})$ is concave in $\overline{\theta_r}$ (Condition 2 in Definition 4), if at least two components of $\boldsymbol{\theta}$ are nonzero then the conditions (EC.6)-(24) are necessary and sufficient for $\boldsymbol{\theta}$ to be a Nash equilibrium. Furthermore, if $\boldsymbol{d} \geq 0, \mu > 0, \theta_r = \mu d_r$, and $\sum_{r=1}^{R} d_r = C$, then it follows that $\mu = \sum_{s=1}^{R} \theta_s/C$, $\Phi(\mu) = p_D(\boldsymbol{\theta})$, and $d_r = D(p_D(\boldsymbol{\theta}), \theta_r)$. Thus the conditions (EC.6)-(24) become equivalent to (14)-(15), as required.

*Steps 7, Proof of Theorem 6: Let (C,R,**U**) be a utility system. Then there exists a unique Nash equilibrium.* Our approach will be to demonstrate existence of a Nash equilibrium by finding a solution $\mu > 0$ and $\boldsymbol{d} \geq 0$ to (14)-(15), such that $\sum_{r=1}^{R} d_r = C$. If we find such a solution, then at least two components of $\boldsymbol{d}$ must be nonzero; otherwise, (14) cannot hole for the user r with $d_r = C$. If we define $\boldsymbol{\theta} = \mu d$, then $\mu = \sum_{s=1}^{R} \theta_s/C$, so $p_D(\boldsymbol{\theta}) = \Phi(\mu)$; and from (11), we have $d_r = D(p_D(\boldsymbol{\theta}), \theta_r)$. Thus if $\mu > 0$ and $\boldsymbol{d} \geq 0$ satisfy (14)-(15), then $\boldsymbol{\theta} = \mu d$ is a Nash equilibrium by Steps 6. Consequently, it suffices to find a solution $\mu > 0$ and $\boldsymbol{d} \geq 0$ to (14)-(15).

We first show that for a fixed value of $\mu > 0$, the equality in (14) has at most one solution d_r. To see this, rewrite (14) as:

$$U_r'(d_r)\left(1 - \frac{d_r}{C}\right) - (\mu\Phi'(\mu) - \Phi(\mu))\left(\frac{d_r}{C}\right) = \Phi(\mu). \tag{25}$$

Since Φ is convex and strictly increasing with $\Phi(\mu) \to 0$ as $\mu \to 0$, we have $\mu\Phi'(\mu) - \Phi(\mu) \geq 0$. Thus the left hand side is strictly decreasing in d_r (since U_r is strictly increasing and concave), from $U_r'(0)$ at $d_r = 0$ to $\mu\Phi'(\mu) - \Phi(\mu) \leq 0$ when $d_r = C$. This implies a unique solution $d_r \in [0, C]$ exists for the equality in (14) as long as $U_r'(0) \geq \Phi(\mu)$; we denote this solution $d_r(\mu)$. If $\Phi(\mu) > U_r'(0)$, then we let $d_r(\mu) = 0$. Observe that as $\mu \to 0$, we must have $d_r(\mu) \to C$, since otherwise we can show that (14) fails to hold for sufficiently small μ.

Next we show that $d_r(\mu)$ is continuous. Since we defined $d_r(\mu) = 0$ if $\Phi(\mu) > U_r'(0)$, and $d_r(\mu) = 0$ if $\Phi(\mu) = U_r'(0)$ from (14), it suffices to show that $d_r(\mu)$ is continuous for μ such that $\Phi(\mu) \leq U_r'(0)$. But in this case continuity of d_r can be shown using (14), together with the fact that U_r', Φ and Φ' are all continuous (the latter because Φ is concave and differentiable, and hence continuously differentiable). Indeed, suppose that $\mu_n \to \mu$ where $\Phi(\mu) \leq U_r'(0)$, and assume without loss of generality that $d_r(\mu_n) \to d_r$ (since $d_r(\mu_n)$ takes values in the compact set [0,C]). Then since μ_n and $d_r(\mu_n)$ satisfy the equality in (14) for sufficiently large n, by taking limits we see that μ and d_r satisfy the equality in (14) as well. Thus we must have $d_r = d_r(\mu)$, so we conclude $d_r(\mu)$ is continuous.

We now show that $d_r(\mu)$ is nonincreasing in μ. To see this, choose $\mu_1, \mu_2 > 0$ such that $\mu_1 < \mu_2$. Suppose that $d_r(\mu_1) < d_r(\mu_2)$. Then, in particular, $d_r(\mu_2) > 0$, so (14) holds with equality for $d_r(\mu_2)$ and μ_2. Now note that as we move from $d_r(\mu_2)$ to $d_r(\mu_1)$, the left hand side of (14) strictly increases (since U_r is concave). On the other hand, since Φ is convex and strictly increases with $\Phi(\mu) \to 0$ as $\mu \to 0$, we have the inequalities $\mu_2\Phi'(\mu_2) - \Phi(\mu_2) \geq \mu_1\Phi'(\mu_1) - \Phi(\mu_1) \geq 0$. From this it follows that the right hand side of (14) strictly decreases as we move from $d_r(\mu_2)$ to $d_r(\mu_1)$ and from μ_2 to μ_1. Thus neither (14) nor (15) can hold at $d_r(\mu_1)$ and μ_1; so we conclude that for all r, we must have $d_r(\mu_1) \geq d_r(\mu_2)$.

Thus for each r, $d_r(\mu)$ is a nonincreasing continuous function such that $d_r(\mu) \to C$ as $\mu \to 0$, and $d_r(\mu) \to 0$ as $\mu \to \infty$. We conclude there exists at least one $\mu > 0$ such that $\sum_{r=1}^{R} d_r(\mu) = C$; and in this case $\boldsymbol{d}(\mu)$ satisfies (14)-(15), so by the discussion at the beginning of this step, we know that $\boldsymbol{\theta} = \mu\boldsymbol{d}(\mu)$ is a Nash equilibrium.

Finally, we show that the Nash equilibrium is unique. Suppose that there exist two solutions $\boldsymbol{d}^1 \geq 0, \mu_1 > 0$, and $\boldsymbol{d}^2 \geq 0, \mu_2 > 0$ to (14)-(15), such that $\sum_{r=1}^{R} d_r^i = C$ for $i = 1, 2$. Of course, we must have $\boldsymbol{d}^i = \boldsymbol{d}(\mu_i), i = 1, 2$. We assume without loss of generality that $\mu_1 \leq \mu_2$; our goal is to show that $\mu_1 = \mu_2$. Since $d_r(\cdot)$ is nonincreasing, we know $d_r(\mu_1) \geq d_r(\mu_2)$ for all r. Since $\sum_{r=1}^{R} d_r^i = C$ for $i = 1, 2$, we conclude that $d_r(\mu_1) = d_r(\mu_2)$ for every r. Let r be such that $d_r(\mu_1) = d_r(\mu_2) > 0$. Observe that $\Phi(\mu)$ and $\mu\Phi'(\mu)$ are both strictly increasing in $\mu > 0$, since Φ is strictly increasing and convex. Thus for fixed $d_r > 0$, the equality in (14) has a unique solution μ, so $d_r(\mu_1) = d_r(\mu_2) > 0$ implies $\mu_1 = \mu_2$. Thus (14)-(15) have a unique solution $\boldsymbol{d} \geq 0, \mu > 0$, such that $\sum_{r=1}^{R} d_r = C$. From Step 6, this ensures the Nash equilibrium $\boldsymbol{\theta} = \mu\boldsymbol{d}$ is unique as well. Thus, combining steps 1 to 7, we prove Theorem 6. ■

ACKNOWLEDGEMENT

The authors would like to thank Prof. Edgar Whitley of the London School of Economics and Political Science for his insightful comments on the market methodology.

CONTRIBUTION STATEMENT

Ranjan Pal and Jon Crowcroft designed the market methodology. Ranjan Pal, Yixuan Wang, Mingyan Liu, Swades De, and Bodhibrata Nag analyzed the market settings. Yong Li, Sasu Tarkoma, Abhishek Kumar, and Pan Hui helped gather experimental data, and run computer simulations.

REFERENCES

[1] A. Acquisti, C. Taylor, and L. Wagman, "The economics of privacy," *J. Econ. Literature*, vol. 54, no. 2, pp. 92–442, 2016.

[2] V. Benndorf and H. Normann, "The willingness to sell personal data," *Scandin. J. Econ.*, vol. 120, no. 4, pp. 1260–1278, Oct. 2018.

[3] N. Lomas, "The case against behavioral advertising is stacking up," *TechCrunch*, Jan. 2019. [Online]. Available: https://techcrunch.com/2019/01/20/dont-be-creepy/

[4] S. B. Barnes, "A privacy paradox: Social networking in the United States," *First Monday*, vol. 11, no. 9, Sep. 2006. [Online]. Available: https://firstmonday.org/ojs/index.php/fm/article/view/1394, doi: 10.5210/fm.v11i9.1394.

[5] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London, U.K.: Profile Books, 2019.

[6] R. Pal and J. Crowcroft, "Privacy trading in the surveillance capitalism age viewpoints on 'privacy-preserving' societal value creation," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 49, no. 3, pp. 26–31, Nov. 2019.

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

[8] Wikipedia. (2018). *Facebook-Cambridge Analytica Data Scandal*. [Online]. Available: https://en.wikipedia.org/wiki/Facebook-Cambridge_Analytica_data_scandal

[9] P. D. Klemperer and M. A. Meyer, "Supply function equilibria in oligopoly under uncertainty," *Econometrica, J. Econ. Soc.*, pp. 1243–1277, 1989.

[10] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–38, Jul. 2018.

[11] R. Poser, "The right of privacy," *Georgia Law Rev.*, vol. 12, no. 3, p. 393, 1978.

[12] R. Poser, "The economics of privacy," *Amer. Econ. Rev.*, vol. 71, no. 2, pp. 405–409, 1981.

[13] G. J. Stigler, "An introduction to privacy in economics and politics," *J. Legal Stud.*, vol. 9, no. 4, pp. 623–644, Dec. 1980.

[14] C. Kenneth Laudon, "Markets and privacy," *Commun. ACM*, vol. 39, no. 9, pp. 92–104, Sep. 1996.

[15] H. R. Varian, "Economic aspects of personal privacy," in *Internet Policy and Economics*, W. Lehr and L. Pupillo, Eds. Boston, MA, USA: Springer, 2009, doi: 10.1007/b104899_7.

[16] A. Odlyzko, "Privacy, economics, and price discrimination on the Internet," in *Economics of Information Security* (Advances in Information Security), vol. 12, L. J. Camp and S. Lewis, Eds. Boston, MA, USA: Springer, 2003, doi: 10.1007/1-4020-8090-5_15.

[17] P. M. Schwartz, "Property, privacy, and personal data," *Harvard Law Rev.*, vol. 117, no. 7, p. 2056, May 2004.

[18] P. Samuelson, "Privacy as intellectual property," *Stanford Law Rev.*, vol. 52, pp. 1125–1173, May 2000.

[19] J. Hirschleifer, "The private and social value of information and the reward to inventive activity," *Amer. Econ. Rev.*, vol. 61, no. 4, 1971.

[20] J. Hirshleifer, "Privacy: Its origin, function, and future," *J. Legal Stud.*, vol. 9, no. 4, pp. 649–664, Dec. 1980.

[21] J. M. Burke, C. R. Taylor, and L. Wagman, "Information acquisition in competitive markets: An application to the US mortgage market," *Amer. Econ. J., Microecon.*, vol. 4, no. 4, pp. 65–106, Nov. 2012.

[22] L. Wagman, "Good news or bad news?: Information acquisition and applicant screening in competitive labor markets," *SSRN*, Jan. 2014. [Online]. Available: https://ssrn.com/abstract=2377849, doi: 10.2139/ssrn.2377849.

[23] A. F. Daughety and J. F. Reinganum, "Public goods, social pressure, and the choice between privacy and publicity," *Amer. Econ. J., Microecon.*, vol. 2, no. 2, pp. 191–221, May 2010.

[24] M. Spence, "Job market signaling," *Quart. J. Econ.*, vol. 87, no. 3, pp. 355–374, 1973.

[25] R. H. Coase, "The problem of social cost," in *Classic Papers in Natural Resource Economics*. London, U.K.: Palgrave Macmillan, 1960, pp. 87–137.

[26] P. Bolton and M. Dewatripont, *Contract Theory*. Cambridge, MA, USA: MIT Press, 2005.

[27] P. Jain, M. Gyanchandani, and N. Khare, "Enhanced secured map reduce layer for big data privacy and security," *J. Big Data*, vol. 6, no. 1, pp. 1–17, Dec. 2019.

[28] S. Madan and P. Goswami, "A privacy preserving scheme for big data publishing in the cloud using k-Anonymization and hybridized optimization algorithm," in *Proc. Int. Conf. Circuits Syst. Digit. Enterprise Technol. (ICCSDET)*, Dec. 2018, pp. 1–7.

[29] C. Thota, R. Sundarasekar, G. Manogaran, R. Varatharajan, and M. Priyan, "Centralized fog computing security platform for iot and cloud in healthcare system," in *Fog Computing, Breakthroughs in Research and Practice*. Hershey, PA, USA: IGI Global, 2018, pp. 365–378.

[30] Z. A. Al-Odat and S. U. Khan, "Anonymous privacy-preserving scheme for big data over the cloud," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 5711–5717.

[31] A. De Corniere and R. De Nijs, "Online advertising and privacy," *RAND J. Econ.*, vol. 47, no. 1, pp. 48–72, 2016.

[32] J. Levin and P. Milgrom, "Online advertising: Heterogeneity and conflation in market design," *Amer. Econ. Rev.*, vol. 100, no. 2, pp. 603–607, May 2010.

[33] D. Bergemann and A. Bonatti, "Targeting in advertising markets: Implications for offline versus online media," *RAND J. Econ.*, vol. 42, no. 3, pp. 417–443, Sep. 2011.

[34] S. Cowan, "The welfare effects of third-degree price discrimination with nonlinear demand functions," *RAND J. Econ.*, vol. 38, no. 2, pp. 419–428, Jun. 2007.

[35] A. Ghosh and A. Roth, "Selling privacy at auction," *Games Econ. Behav.*, vol. 91, pp. 334–346, May 2015.

[36] L. K. Fleischer and Y.-H. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data," in *Proc. 13th ACM Conf. Electron. Commerce EC*, 2012, pp. 568–585.

[37] K. Ligett and A. Roth, "Take it or leave it: Running a survey when privacy comes at a cost," in *Proc. Int. Workshop Internet Netw. Econ.* Berlin, Germany: Springer, 2012, pp. 378–391.

[38] A. Roth and G. Schoenebeck, "Conducting truthful surveys, cheaply," in *Proc. 13th ACM Conf. Electron. Commerce EC*, 2012, pp. 826–843.

[39] A. Ghosh and K. Ligett, "Privacy and coordination: Computing on databases with endogenous participation," in *Proc. 14th ACM Conf. Electron. Commerce*, Jun. 2013, pp. 543–560.

[40] K. Nissim, S. Vadhan, and D. Xiao, "Redrawing the boundaries on purchasing data from privacy-sensitive individuals," in *Proc. 5th Conf. Innov. Theor. Comput. Sci. ITCS*, 2014, pp. 411–422.

[41] A. Ghosh, K. Ligett, A. Roth, and G. Schoenebeck, "Buying private data without verification," in *Proc. 15th ACM Conf. Econ. Comput. EC*, 2014, pp. 931–948.

[42] D. Xiao, "Is privacy compatible with truthfulness?" in *Proc. 4th Conf. Innov. Theor. Comput. Sci. ITCS*, 2013, pp. 67–86.

[43] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan, "Truthful mechanisms for agents that value privacy," *ACM Trans. Econ. Comput.*, vol. 4, no. 3, pp. 1–30, Jun. 2016.

[44] W. Wang, L. Ying, and J. Zhang, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 44, no. 1, pp. 249–260, Ju. 2016.

[45] M. M. Khalili, X. Zhang, and M. Liu, "Contract design for purchasing private data using a biased differentially private algorithm," in *Proc. 14th Workshop Econ. Netw., Syst. Comput. NetEcon*, 2019, pp. 1–6.

[46] 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 *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Apr. 2019, pp. 1045–1053.

[47] C. Dwork, "Differential privacy," in *Automata, Languages and Programming*. Berlin, Germany: Springer-Verlag, 2006, pp. 1–12.

[48] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[49] T. Roughgarden, *Selfish Routing Price Anarchy*, vol. 174. Cambridge, MA, USA: MIT Press, 2005.

[50] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*, vol. 1. New York, NY, USA: Oxford Univ. Press, 1995.

[51] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, vol. 23. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[52] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[53] H. R. Varian, "Microeconomic analysis," WW Norton, New York, NY, USA, Tech. Rep., 1992, p. 506. [Online]. Available: http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=orton.xis&method=post&formato=2&cantidad=1&expresion=mfn=070755

[54] K. J. Arrow and M. D. Intriligator, *Handbook of Mathematical Economics*, vols. 1–3. Amsterdam, The Netherlands: North Holland, 1981. [Online]. Available: http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=orton.xis&method=post&formato=2&cantidad=1&expresion=mfn=079507

[55] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[56] R. Johari and J. N. Tsitsiklis, "Efficiency of scalar-parameterized mechanisms," *Oper. Res.*, vol. 57, no. 4, pp. 823–839, Aug. 2009.

RANJAN PAL (Member, IEEE) received the Ph.D. degree in computer science from the USC's Viterbi School of Engineering. He was a Postdoctoral Fellow with the CS Department, and the Center for Mathematical Sciences, University of Cambridge. He is currently a junior Faculty Member of ECE with the University of Michigan, Ann Arbor, MI, USA. His research interests include engineering robust cyber-security and information privacy solutions using decision and the applied mathematical sciences. He also moonlights in the economics of networked and distributed information systems. He is a member of ACM, AMS, INFORMS, SIAM, and the Game Theory Society. He serves as an Editor of the IEEE Networking Letters and the *ACM Transactions on Management Information Systems*.

JON CROWCROFT (Fellow, IEEE) is currently the Marconi Professor of communications systems with the Computer Laboratory, University of Cambridge, and the Director of the Alan Turing Institute, London. His current active research areas are opportunistic communications, social networks, privacy preserving analytics, and techniques and algorithms to scale infrastructure-free mobile systems. He is a Fellow of the Royal Society, the Royal Academy of Engineering, the ACM, and the British Computer Society. He was a recipient of the ACM SIGCOMM Lifetime Achievement Award and the IEEE Internet Award.

YIXUAN WANG (Student Member, IEEE) is currently pursuing the bachelor's degree in computer science with the University of Michigan, Ann Arbor, MI, USA. Her main research interests include machine learning, privacy, and human–computer interaction. She is a Student Member of the ACM.

YONG LI (Senior Member, IEEE) received the Ph.D. degree in electronics engineering from Tsinghua University, Beijing, China, in 2012. He was a Postdoctoral Fellow with the Telekom Labs, Germany. He is currently a Faculty Member with the Department of Electronic Engineering, Tsinghua University. His research interests are in the areas of big data, mobile computing, wireless communications, and networking. He serves as an Editor of the IEEE Transactions on Network and Service Management, and ACM IMWUT.

SWADES DE (Senior Member, IEEE) is currently a Professor of electrical engineering with the Indian Institute of Technology Delhi. His current directions include energy harvesting communications, broadband wireless access, cognitive/white-space access networks, smart grid, and IoT communications. He is a Fellow of the Indian National Academy of Engineering. He serves as an Editor of the IEEE Networking Letters, the IEEE Wireless Communication Letters, and the IEEE Transactions on Vehicular Technology.

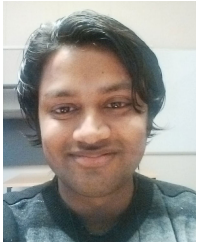
SASU TARKOMA (Senior Member, IEEE) is currently a Professor of computer science with the University of Helsinki, the Head of the Department of Computer Science, the Director of the Nokia Center for Advanced Research, and the Director of the Helsinki Center for Data Science. He has seven granted U.S. patents. His research interests are Internet technology, distributed systems, and mobile and ubiquitous computing. He was a recipient of several best paper awards, e.g., the IEEE PerCom, ACM SIGCOMM Computer Communication Review, and ACM SIGOPS Operating System Review.

MINGYAN LIU (Fellow, IEEE) is currently an Entrepreneur and the Peter and Evelyn and the Fuss Chair Professor of electrical and computer engineering with the University of Michigan, Ann Arbor, MI, USA. Her research interests are in optimal resource allocation, sequential decision theory, incentive design, online learning, and modeling and mining of large scale Internet measurement data concerning cyber security. She was a co-founder of the cybersecurity scoring startup Quadmetrics, in 2014, that was acquired by FICO, in 2016. She is an ACM Member.

BODHIBRATA NAG (Senior Member, IEEE) received the Ph.D. degree in operations research and system analysis from the Indian Institute of Management (IIM), Calcutta.

He is currently the Dean (Academic) and a Professor with the Operations Management Group, IIM. His research interests are the application of analytic and optimization techniques for planning, design and operations of energy, and logistics systems.

He is a Senior Member of INFORMS.



ABHISHEK KUMAR (Student Member, IEEE) received the M.S. degree in knowledge service engineering from KAIST, in 2016. He is currently pursuing the Ph.D. degree in computer science with the University of Helsinki.

His research interests broadly include security and privacy in Apps and the IoT space, human–computer interaction, and applied machine learning. He is a Student Member of the ACM.

PAN HUI (Fellow, IEEE) was an Adjunct Professor in social computing and network with Aalto University, from 2012 to 2016, a Senior Research Scientist and then a Distinguished Scientist for Telekom Innovation Laboratories (T-Labs), Germany, from 2008 to 2015. He is currently a Nokia Chair Professor of computer science with the University of Helsinki, Finland, and a Professor with the Department of Computer Science and Engineering, HKUST, Hong Kong. His current research interest include the applications of data science to design modern social and communication networks. He is an ACM Distinguished Scientist and a member of the Academy of Europe.