Technical Report

UCAM-CL-TR-925
ISSN 1476-2986

Number 925

UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

Privacy markets in the Apps and IoT age

Ranjan Pal, Jon Crowcroft, Abhishek Kumar,
Pan Hui, Hamed Haddadi, Swades De,
Irene Ng, Sasu Tarkoma, Richard Mortier

September 2018

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

https://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

https://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

Privacy Markets in the Apps and IoT Age

Ranjan Pal*, Jon Crowcroft, Abhishek Kumar, Pan Hui, Hamed Haddadi, Swades De, Irene Ng, Sasu Tarkoma, Richard Mortier

Abstract

In the era of the mobile apps and IoT, huge quantities of data about individuals and their activities offer a wave of opportunities for economic and societal value creation. However, the current personal data ecosystem is fragmented and inefficient. On one hand, end-users are not able to control access (either technologically, by policy, or psychologically) to their personal data which results in issues related to privacy, personal data ownership, transparency, and value distribution. On the other hand, this puts the burden of managing and protecting user data on apps and ad-driven entities (e.g., an ad-network) at a cost of trust and regulatory accountability. In such a context, data holders (e.g., apps) may take advantage of the individuals' inability to fully comprehend and anticipate the potential uses of their private information with detrimental effects for aggregate social welfare. In this paper, we[1] investigate the problem of the existence and design of efficient ecosystems (modeled as markets in this paper) that aim to achieve a maximum social welfare state amongst competing data holders by preserving the *heterogeneous* privacy preservation constraints upto certain compromise levels, induced by their clients, and at the same time satisfying requirements of agencies (e.g., advertisers) that collect and trade client data for the purpose of targeted advertising, assuming the potential practical inevitability of some amount inappropriate data leakage on behalf of the data holders. Using concepts from supply-function economics, we propose the first mathematically rigorous and provably optimal privacy market design paradigm that always results in unique equilibrium (i.e, stable) market states that can be either economically efficient or inefficient, depending on whether privacy trading markets are monopolistic or oligopolistic in nature. Subsequently, we characterize in closed form, the efficiency gap (if any) at market equilibrium.

1 Introduction

Mobile applications (apps) and the IoT are driving the modern digital ecosystem. In-app advertising is an essential part of the ecosystem of free mobile applications. On the surface, this creates a win-win situation where app developers can profit from their work without charging the users. Meanwhile, ad networks employ personalization to improve the effectiveness/profitability of their ad placement. This need for serving personalized advertisements in turn motivates ad networks to collect profile data about users. As such, "free" apps are only free in monetary terms; they come with the price of potential

*Corresponding Author, Email: rp631@cam.ac.uk, rpal@usc.edu

[1]R. Pal, J. Crowcroft, and R. Mortier are with University of Cambridge, UK; A. Kumar, P. Hui, and S. Tarkoma are with University of Helsinki, Finland; H. Haddadi is with Imperial College London, UK; S. De is with Indian Institute of Technology Delhi, India; I. Ng is with University of Warwick, UK, and is the current CEO of HATDEX.

privacy concerns. In comparison to in-browser advertising, research focused on mobile ad personalization is a significant pursuit for the following reasons [1]: 1) Mobile devices are a lot more intimate to users; they are carried around at all times and are being used more and more for sensitive operations like personal communications, dating, banking, etc. Therefore, privacy concerns regarding what information is collected for ad personalization are more serious. 2) Unlike in-browser advertising, where the advertisement content is strictly isolated from the rest of the displayed page by the well-known "same origin policy", in-app advertising operates in a new and less understood environment. Like in the mobile app ecosystem, IoT has the potential to provide enormous benefits for consumers, but it also has significant privacy and security implications arising due to commercial requirements of data generated in such systems. The IoT could improve global health, modernize city infrastructures, and spur global economic growth. To be sure, these potential benefits are immense, but so too are the potential risks: Connected devices that provide increased convenience and improve health services also collect, transmit, store, and often share vast amounts of consumer data; some of it highly personal, thereby creating privacy risks.

1.1 Research Motivation

Privacy risks in the current mobile app and IoT age has led to the personal data ecosystem to be fragmented and inefficient. On one hand end-users are less empowered to control access (either technologically, by policy, or psychologically) to their personal data which results in issues related to privacy, personal data owner ship, transparency, and value distribution. On the other hand it puts the burden of managing and protecting user data on apps and ad-driven entities (e.g., an ad-network) at a cost of trust and regulatory accountability. In such a context, data holders (e.g., apps) may take advantage of the individuals inability to fully comprehend and anticipate the potential uses of their private information with detrimental effects for aggregate social welfare. As a well known example of this context, the recent *Facebook-Cambridge Analytica* data scandal [2] confirmed this aforementioned inability of individuals. Personal information of Facebook users were obtained originally through a Facebook (Personality Test Quiz) application, thisisyourdigitallife, with users' consent. Later on, the information was shared with Cambridge Analytica without users' consent, who in turn used this data to influence voter opinion during 2017 US presidential election on behalf of politicians who hired them.

As an example of a step to mitigate privacy risks, the European Union has recently introduced new General Data Protection Regulation (GDPR), which came into effect in May 2018, and is explicitly concerned to handle the threat to privacy occasioned by the emerging digital ecosystem [3]. As a key challenge, GDPR seeks to put in place measures to address an accountability requirement. Accountability requires that any organization controlling data processing put in place policies, procedures and systems to demonstrate to itself that its processing operations comply with the requirements of data protection regulation. Equally important, is the "external" dimension of accountability, which requires that a data processing entity demonstrate to others, particularly regulatory authorities and individual data subjects, that its data processing operations comply with regulation (e.g., Cambridge Analytica will need to delete user data on the latter's request in the GDPR regime).

Despite a regulation like the GDPR that is being currently put into place in some parts of the world, the personal data ecosystem that is likely to be dominated by the mobile and IoT industry in the near future, might remain fragmented and inefficient to a certain degree due to three primary reasons: (a) regulations such as GDPR are yet to be

pervasive throughout the world, and most governments are slow to come to terms with them. E.g., Currently the USA does not have such regulations and companies are taking full advantage of this. Recently, Bloomberg reported a secret deal between Google and Mastercard in which Google advertisers were provided tools to track whether the ads they ran online led to a sale at a physical store in the US. Most of the two billion Mastercard users do not know about this behind-scene tracking [4], (b) unlike in the browser-space, in the mobile space, technologies such as ad-blockers that give users the power to control the show of potentially privacy hampering advertisements, are to a great extent ineffective due to usability reasons [5][6], and (c) the desperate mindset of advertisers to make consumer data be of commercial interest to them (for the purpose of targeted advertising as an example) will at best make way for the design of voluntary (and in some cases controlled) consumer data releasing mechanisms (e.g., via the IoT hub [3][7]), that on one side will try to ensure the preservation of consumer privacy, but on the other hand will leave open cracks in the design that will contribute to privacy risks (e.g., via social engineering attacks launched by taking advantage of human psychological aspects). *Thus, a significant challenge is the design of mechanisms that accept the inevitability of unwanted consumer data release and minimize privacy risks at the same time.*

Research Goal - Our goal in this paper is to design and analyze an efficient mechanism for minimizing privacy risk for data release environments in the mobile and IoT space that mutually satisfies the interests of various stakeholders involved in the data release process, viz., consumers, competing data holders (e.g., ad-publishing apps), ad-networks, and advertisers.

1.2 Research Contributions

We make the following research contributions in this paper. Definitions of basic terms in economics are briefly explained in the Appendix.

- We model the aforementioned stakeholder ecosystem setting as a supply-demand market consisting of consumers, competing (both, in a perfect and also in an oligopolistic sense) data holders with locked-in consumer base, ad-networks, and advertisers. We coin this market model as *Privacy Bazaar*. A salient feature of Privacy Bazaar is the use of data holder *supply functions* [8] that characterize the amount of privacy compromise each data holder is willing to make, i.e., the supply, for a given "benefit" it receives from the ad-network. The data holders submit as bids - their supply functions to the ad-network, which then executes a uniform market clearing "benefit" mechanism for all competing data holders, that is aimed at achieving optimal utilitarian social welfare at market equilibria - an efficient state where all stakeholders are mutually optimally satisfied (see Section: 2). For reasons to be made clear in Section: 2, we will use parameterized versions of supply functions as introduced in [9].

- As an ideal benchmarking task, we first investigate a perfectly competitive market and show that it achieves a maximum utilitarian social welfare state at a unique competitive equilibrium. We then investigate oligopolistic markets in which due to strategic "benefit" anticipating behavior of the data holders, the market equilibrium is less efficient than in the perfectly competitive scenario, where data holders are "benefit" taking. In this regard, we mathematically characterize the efficiency loss 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. Specifically, we find the following:(a) the set of data-holders

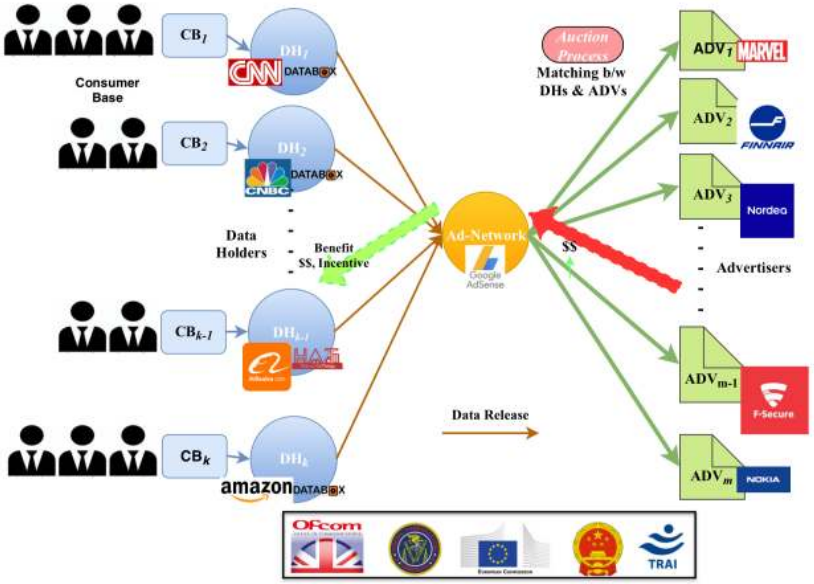


Figure 1: Illustration of a Privacy Trading Market Architecture

at Oligopolistic Nash equilibrium (ONE) who compromise on their privacy requirements, 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 (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, the differences between the PCE and ONE tend to be very small - if the cost functions are extremely heterogeneous, the quantification of the differences can serve as rules of thumb for the ad-network to limit the compromising power of large data holder firms to promote utilitarian social welfare (see Section: 3).

- We extend the above scenario to investigate perfectly competitive and oligopolistic markets for the case when data holders have a lower and upper bound on their privacy compromise amount. Not surprisingly, we show that perfectly competitive markets achieve a maximum utilitarian social welfare state at a unique equilibrium. For oligopoly markets, we show that ONE is unique and is less efficient compared to PCE in regard to maximum utilitarian social welfare, but the efficiency loss is upper bounded, and is non-decreasing in the largest (in terms of compromise capacity) data holder's compromise constraint, and is strictly decreasing in the total compromise of other data holders (see Section: 4).

- We design efficient and scalable distributed supply function bidding algorithms that converge to market equilibria in both perfectly competitive as well as oligopolistic settings (see Section: 5).

2 System Model

In this section, we propose the salient features of *Privacy Bazaar*, our market model based on the parameterized version of the seminal economic theory of supply function bidding proposed by Klemperer in [8]. Table 1 can be referred to for a set of important notations used in the paper.

2.1 Market Elements

Our market elements comprise of **consumers**, **data holders** (DHs), an **ad-network**, and **advertisers** (ADVs) (see Figure 1).

We assume that consumers are locked-in with their respective data holders. Examples of data holders include ad-publishing mobile apps, social media, promotional emails, and IoT databoxes[2]. 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 [10]. 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. This is done by the ad-network usually through an algorithmic matching process [11] - the design of which is not the focus of our paper.

2.2 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[3] using a supply function bidding process (see below). 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 ADVs pay the ad-network to match them with appropriate DHs so as to enable targeting, tracking, and reporting of consumer impressions. This is usually done through a bidding process like VickreyClarkeGroves (VCG) auction (not the explicit focus of this work - see [11] for details) between the ADVs and the ad-network, based on consumer data that interests relevant ADVs.

2.3 (Parameterized) Supply Function Bidding

In this section we propose the supply function bidding process between competing DHs and the ad-network for the case when DHs have no constraints on their privacy compromise amount.

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. In this work, 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* [12] that measure the amount of information an adversary can gain. 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*

[2]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.

[3]The case of competing ad-networks will be our future work.

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 [13] 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 compromise $q_i(b_i, p_i)$ amounts of privacy (measured through the privacy metric) with respect to its consumer base data with the ad-network, in return for a benefit, p_i, i.e., q_i is a parameterized function of p_i and a non-negative bidding parameter b_i. In this work we will assume q_i to be a linear function (rationale explained below) of the form:

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

The compromise function, q_i, for each DH i is their *parameterized supply function*. Examples of benefits include the amount of price *reduction* over the market price paid by individual consumers locked-in with a given DH[4], 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 experience. However, since the heterogeneous revenue functions for individual DHs are private information, the privacy compromise amount, q_i for which each DH would prefer the same benefit cleared by the ad-network, is different for the different DHs. 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. The ad-network just has the values of q_i's at its disposal to arrive at a market clearing value that maximizes utilitarian social welfare amongst the DHs. We assume that the total privacy compromise needs to meet a specific amount $d > 0$ for the ad-network when it clears the market, i.e.,

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

or

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

Here, $b = (b_1, ..., b_N)$ is the supply function profile of the DHs. In the event when $\sum_i b_i = 0$, the ad-network will reject the bid. Please note here that no DH has any bounds, i.e., not constrained on their privacy compromise amount. In reality, each DH might have upper and lower bounds on their privacy compromise amount, and we will deal with this case as part of future work.

2.4 Framework Justification

In this section, we justify the use of supply function bidding framework in light of the following questions:

Why Use Supply Function Bidding Mechanisms? - Supply function as a strategic variable allows to adapt better to changing market conditions (such as uncertain, variable, or stochastic supply of privacy compromise by DHs) than does a simple commitment to a fixed compromise quantity [8], because no matter what the value of the supply deficit is, the utility company can use the supply function bid by the customers to clear the deficit. The other motivation to use supply function is to respect practical informational

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

constraints in the market ecosystem. A DH might not want to reveal its cost function because of incentive or security concerns which means more communication. A properly-chosen parameterized supply function controls information revelation while requiring less communication.

Why Use a Linear Supply Function? The seminal work in [8] used the general function as the bidding strategy. When the bidding action is changed from the linear form (represented by the single variable, b_i in our work) to a general form, the analysis of the strategic behavior of the DHs become much more complicated. To solve the general supply function equilibrium (SFE) (introduced in [8]) 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. For practical applications, the asymmetric case is more interesting. 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 paper, 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.

Why Use Supply Function Bidding and Not VCG Auctions? - A well known technique to approach our market design problem is to use the seminal VCG class of mechanisms, which entail truthful reporting of compromise preferences by individual DHs, as a dominant strategy that maximizes utilitarian social welfare. However, there are several reasons why a VCG mechanism may not be desirable in practical settings. As an example, there is no bound on the benefit the ad-network may have to make to DHs. In addition, VCG mechanisms exhibit the implicit "unfairness" of providing different benefits to different market participants (in our case DHs), something that is not the design choice for our problem. See [14][15][16] for extensive discussion of some of the shortcomings of the VCG mechanism. It is worth noting that several papers have studied approaches to price or allocate "benefits" to sharable resources (e.g., privacy compromise amount) using VCG-like mechanisms with scalar strategy spaces; see, e.g., [17][18]. Similar approaches could be applied in our context to yield efficient or nearly-efficient market mechanisms, though with attendant shortcomings analogous to standard VCG mechanisms.

3 Markets Sans Compromise Bounds

In this section, we analyze perfectly competitive and oligopolistic market structures of DH competition with respect to privacy compromise strategies in the backdrop of a single ad-network, when there are no restrictions on DH compromise amounts.

3.1 Perfectly Competitive Markets

In perfectly competitive markets, DHs are benefit taking. Given a benefit p, each DH i maximizes its net revenue given as:

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

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 cost incurred to make the compromise. This cost can be interpreted as the sum of the amount of cost of technical adjustments required to compromise privacy

Table 1: Table of Important Notations

N	set 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
L_i	lower limit of the compromise amount for DH i
D_i	upper limit of the compromise amount for DH i
π_i	payoff for DH i
LR_i	Lerner's Index of DH i

(e.g., technological costs of hosting ads by advertisers) and cost of handling consumer complaints/unpopularity with respect to degradation of quality of experience (QoE).

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 Appendix.

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 Implication - 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 a 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 $C_1^0 \leq C_2^0 \leq ..... \leq C_{|N|}^0$. For modeling convenience, we also introduce parameter $C_{|N|+1}^0$ and set its value to $C_n^{'}(d)$. Thus, we have $C_1^0 \leq C_2^0 \leq ..... \leq C_{|N|}^0 \leq C_{|N|+1}^0$. We have the following result on the privacy compromise characteristics of individual DHs, the proof of which is in the Appendix.

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* $\bar{N} = \{1, 2, ..., \bar{n}\}$*, 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 Implication - 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 Appendix.

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* $i \in \bar{N}$*.*

3.2 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. We denote the supply function for all DHs but i as $b_{-i} = (b_1, b_2, ..., b_{i-1}, b_{i+1}, ..., b_{|N|})$ 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}

$$u_i(b_i, b_{-i}) = p(b)q_i(p(b), b_i) - C_i(q_i(p(b), b_i)) = \frac{d^2 b_i}{(\sum_j b_j)^2} - C_i\left(\frac{db_i}{(\sum_j b_j)}\right). \tag{7}$$

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*

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

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

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 Appendix. We now turn to state *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 Implication - 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 later in the paper for an 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 Appendix, 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.

Based on the above theorem, 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 $C_1^0 \leq C_2^0 \leq ..... \leq C_{|N|}^0$. 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 $C_1^0 \leq C_2^0 \leq ..... \leq C_{|N|}^0 \leq C_{|N|+1}^0$. We now have the *second* important result for this section, on the privacy compromise characteristics of individual DHs, the proof of which is in the Appendix.

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 $N^* = \{1, 2, ..., n^*\}$, 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 Implication - 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 Appendix.

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* $i \in N^*$.

3.3 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) [19].* 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 Implication - The three conditions in the theorem respectively imply the following:

- The set of DHs that contribute to the privacy compromise program at ONE is a superset of that at PCE (due to the non-strategic nature of the DHs at PCE).
- The benefit at the ONE is higher that 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 and consequently more benefits), 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.

We also see from the theorem that 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 cost. The presence of closed form expressions for the efficiency loss may serve as a guideline to the ad-network to limit the market power for some DHs (in the oligopoly setting) to maximize social welfare (e.g., by allowing the entry of new large DHs in the market to stiffen competition). In addition, 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 are small. The proof of the result is in the Appendix.

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 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, $\frac{p^*}{p} \to \infty$, and $\frac{C^*}{C} \to \infty$. *Thus, we observe that if there exist DHs with extremely heterogeneous cost functions, the efficiency loss of the ONE might be unbounded.*

4 Markets with Compromise Bounds

In this section, we analyze perfectly competitive and oligopolistic DH competition with respect to privacy compromise strategies in the backdrop of a single ad-network, when DHs have upper and lower limits on their privacy compromise amount. As an introductory step, we first extend system model aspects in Section 2.3., i.e., primarily the supply-bidding framework, to mathematically capture compromise limits of DHs, which is then followed by a detailed market analysis.

4.1 System Model

We assume that each DH i has a cost function C_i that maps his privacy compromise amount d_i to his cost of compromise. C_i is continuous and convex with $C_i(0) = 0$, and is strictly increasing over $[0, \infty]$. The convexity of the DH cost functions follows from the usual assumption that DH utility is concave. Like in Section 3, we assume that the total compromise deficit is d, where $d < \sum_{i=1}^{n} D_i$ (to make the compromise deficit problem as explained later, non-trivial), D_i being the upper limit of the privacy compromise amount for each DH i. Likewise L_i is the lower limit of the compromise amount for each DH i.We assume without loss of generality that d, and D_i are positive, and L_i is non-negative.

The ad-network will then solve the following optimization problem:

$$min_q \sum_{i=1}^{n} C_i(q_i)$$

subject to

$$\sum_{i=1}^{n} q_i = d$$

$$L_i \leq q_i \leq D_i, \forall i$$

where q is the vector of privacy compromise amounts for DHs. Alternatively, we would be referring to q and $\{q_i\}_{i=1}^n$ as supply vectors in our paper. An optimal solution, i.e., the minimum aggregate cost, to the above optimization problem will result in a socially optimal allocation, that might be hard to achieve in practice, but will serve as a benchmark solution to compare against practically achieved solutions.

Parameterized Supply Functions - We restrict the set of supply-functions that DHs can choose from to the following parameterized family introduced in [9] to account for compromise limits:

$$q_i = S_i(b_i, p) = D_i - \frac{b_i}{p}, \tag{10}$$

where b_i is the non-negative bid submitted by DH i, $S_i(b_i, p)$ denotes the amount of privacy compromise DH i would like to undertake at per unit compromise benefit $p > 0$. b_i represents DH i's unwillingness to compromise client privacy, due to a fear of loss of revenue and/or clients. $S_i(b_i, p) = q_i$ for each DH i increases with increasing compromise upper limits D_i, decreasing compromise unwillingness b_i, and increasing benefit p.

In order to clear the market, we have

$$\sum_{i=1}^{n} S_i(b_i, p) = \sum_{i=1}^{n} (D_i - \frac{b_i}{p}) = d. \tag{11}$$

The market clearing benefit parameter p is then given by

$$p = \frac{\sum_{i=1}^n b_i}{-d + \sum_{i=1}^n D_i} \geq 0, \tag{12}$$

where the case $\sum_{i=1}^n b_i = 0$ is ruled out (as done in Section 3) by the ad-network, i.e., re-bidding is done.

Given a positive market-clearing benefit parameter p per unit of privacy compromise, and the bid submitted by DH i, its payoff is given by

$$\pi_i(b_i, p) = pS_i(b_i, p) - C_i(S_i(b_i, p)), \tag{13}$$

or

$$\pi_i(b_i, p) = D_i p - b_i - C_i(S_i(b_i, p)).$$

4.2 Market Analysis

We consider two market settings: one where DHs act as benefit takers (a perfectly competitive DH setting), and the other whether they act as benefit anticipators (an oligopolistic DH setting).

4.2.1 Perfect Competition

Given a benefit p per unit of privacy compromise, a benefit-taking DH i maximizes the payoff function in (13) over $b_i \geq 0$. In this regard, a pair of action (bid) vector and benefit, $(\{b_i\}_{i=1}^n, p)$ forms a perfectly competitive equilibrium if $p > 0$ and

$$\pi_i(b_i, p) = \max_{b_i \geq 0} \pi_i(b_i', p), \ \forall i;$$

and

$$\sum_{i=1}^{n} S_i(b_i, p) = d.$$

We have the following theorem regarding the existence of a perfectly competitive equilibrium, the proof of which is in the Appendix.

Theorem 6. *Given C_i is continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_i D_i > d$, and $n > 1$, there exists a perfectly competitive equilibrium. Furthermore, any competitive equilibrium is an optimal solution to the DH optimization problem:*

$$min_q \sum_{i=1}^{n} C_i(q_i)$$

subject to

$$\sum_{i=1}^{n} q_i = d$$
$$L_i \leq q_i \leq D_i,$$

and is therefore socially optimal.

Theorem Implication - The implication of the above theorem is the same as that of Theorem 1, except that social welfare optimality is now achieved under stricter conditions of DH privacy compromise constraints.

4.2.2 Oligopolistic Competition

Here, we consider DHs with market power who bid strategically to maximize their own payoffs, where the payoff for DH i, $\Pi_i = \pi_i(b_i, b_{-i})$, derived through (13), is given by

$$\Pi_i = \begin{cases} \frac{D_i \sum_{j=1}^{n} b_j}{-d+\sum_{j=1}^{n} D_j} - b_i - C_i\Big(D_i - \frac{b_i(-d+\sum_{j=1}^{n} D_j)}{\sum_{j=1}^{n} b_j}\Big), \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if} \sum_{j=1}^{n} b_j > 0 \\ -C_i\Big(\frac{dD_i}{\sum_j D_j}\Big), \qquad\qquad\qquad \text{if} \sum_{j=1}^{n} b_j, \end{cases} \tag{14}$$

where $b_{-i} = (b_1, ..., b_{i-1}, ..., b_n)$, and in the case when $\sum_{j \neq i} b_j = 0$, $\pi_i(b_i, b_{-i})$ is discontinuous at $b_i = 0$. We then have the following lemma stating the necessary condition for the existence of a Nash equilibrium in an oligopolistic market setting (ONE), the proof of which is in the Appendix.

Lemma 4. *Given C_i is continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_{j \neq i} D_j < d$ for some DH i, and $n > 1$, an ONE does not exist.*

Lemma Implication - The lemma implies that in the condition when $\sum_{j \neq i} D_j < d$, the privacy compromising oligopolistic market fails to enter an equilibrium state since one DH would have the monopoly power. In this regard, a properly chosen (through a market maker) upper threshold on the maximum market-clearing benefit can ensure the existence of an ONE, even if $\sum_{j \neq i} D_i < d$. As a side comment, following from the lemma, a market equilibrium may also fail to exist if $\sum_{j \neq i} D_j = d$ for some DH i. In this case, there could exist infinitely many ONE, all of which have the same form: $b_j = 0 \; \forall j \neq i$, and $b_i > 0$ is large enough for market clearing benefit, p, to be higher than every DH's marginal cost at D_j. The allocation resulting at such ONEs is unique, where DH i compromises zero units of privacy, and the other DHs compromise and amount equal to their upper bound. At such ONEs, DH i despite showing an uncompromising behavior can control market benefit p. Subsequently, arbitrarily high market benefit can result as market equilibrium leading to a arbitrarily high efficiency loss (see Section 4.3), specifically in the case when DH i incurs the lowest cost amongt all the DHs, and thus lead to a situation of market failure. Thus, $\sum_{j \neq i} D_j > d$ is the condition that guarantees that at every ONE, the benefit parameter, p, is determined by atleast two DHs. This automatically leads us to the following lemma, the proof of which is in the Appendix.

Lemma 5. *Given C_i is continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_{j \neq i} D_j > d$, for every DH i, at ONE, b has atleast two positive components.*

We are now led into our main result, i.e., theorem, related to the existence of an oligopolistic Nash equilibrium for our privacy compromise market when competing DHs have lower and upper bounds on their compromise amount. The proof of the theorem is in the Appendix.

Theorem 7. *Given C_i is continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_{j \neq i} D_j > d$, for every DH i, the following properties hold in relation to a privacy compromise market:*

- *There exists an ONE, which is unique if the cost function for every DH is continuously differentiable.*
- *For any ONE b, the resulting allocation $\{S_i(b_i, p(b))\}_{i=1}^n$ is the unique solution to the following optimization problem:*

$$min_q \sum_{i=1}^{n} \hat{C}_i(q_i)$$

subject to

$$\sum_{i=1}^{n} q_i = d$$

$$L_i \leq q_i \leq D_i, \forall i$$

where $\hat{C}_i = \hat{C}_i(q_i)$ for $q_i > 0$ is expressed as:

$$\hat{C}_i = \Big(1 + \frac{q_i}{-d + \sum_{j \neq i} D_j}\Big) C_i(q_i) - \frac{1}{\sum_{j \neq i} D_j} \int_0^{q_i} C_i(x) dx. \tag{15}$$

Theorem Implication - The theorem implies a strong and useful practical insight: in the ideal but likely case when DH cost functions are continuously differentiable, a unique oligopolistic market equilibrium results that does not result in arbitrary market inefficiency - however, though there might arise multiple ONE in the presence of discontinuous DH cost functions (quite likely to arise in practice), all of them surprisingly lead to the *unique* allocation of privacy compromise amongst DHs that solves the optimization problem in the theorem. The following corollary is an immediate outcome of the theorem, the proof of which is in the Appendix.

Corollary 4. *Given C_i is continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_{j \neq i} D_j > d$, for every DH i, at any ONE, every DH achieves a non-negative payoff, and each DH with a non-zero compromise amount achieves a positive payoff.*

Corollary Implication - The corollary implies the satisfaction of the critical individual rationality (IR) property of DHs in the privacy compromise market, without which DHs have no incentive to participate in the market.

An important question to investigate is how compromise bounds affect the market equilibrium. We would expect an increase in a DH's capacity would raise his compromise levels. *However, we surprisingly see the opposite*, as evident from the following result, the proof of which is in the Appendix.

Corollary 5. *Given C_i is continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_{j \neq i} D_j > d$, for every DH i. Also let $\{q_j\}_{j=1}^n$ and $\{\bar{q}_j\}_{j=1}^n$ denote the supply vectors resulting from two ONEs under compromise upper threshold vectors $\{D_j\}_{j=1}^n$ and $\{\bar{D}_j\}_{j=1}^n$ respectively. Let $\bar{D}_j = D_i$, for all $j \neq i$. If $q_i < D_i$ and $\bar{D}_i > D_i$, then $q_i \geq \bar{q}_i$.*

Corollary Implication - Note from (15) that the increase in any DH's compromise upper threshold keeps his modified cost function $\hat{C}$ (see (15)) unchanged, while reducing the $\hat{C}$'s for other DHs. Subsequently, at any ONE where the aggregate cost is minimized, DHs other than i tend to increase their compromise amount which in turn will decrease the compromise amount of DH i.

4.3 Market Efficiency Analysis

Due to strategic competition among firms, market inefficiency is a common phenomenon at ONE in many practical markets. In this section we investigate whether there is loss in market efficiency at ONE for privacy compromise markets, where market efficiency is a state achieved when social welfare, i.e., the negative of sum of costs of DHs, is maximized. We have the following theorem in this regard, the proof of which is in the Appendix.

Theorem 8. *Let C_i be continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_{j \neq i} D_j > d$, for every DH i. Assume*[5] *that there exists a socially optimal allocation vector of privacy compromise among DHs in which every component is non-negative, i.e., $\exists q^*$ that is an optimal solution to the ad-network optimization problem in Section 4.1, such that $q_i^* \geq 0$ for every i. Also let there be a particular i denoting the DH who has the largest compromise limit (threshold). Then for any q at ONE, we have*

$$\sum_{j=1}^{n} C_j(q_j) \leq \Big(1 + \frac{\min\{D_i, d\}}{-d + \sum_{j \neq i} D_j}\Big) \sum_{j=1}^{n} C_j(q_j^*).$$

This bound is tight when $D_i > d$, i.e., for any $\epsilon > 0$, $n \geq 2$, and $D_i \geq d > 0$, there exists $\{D_1, ...D_{i-1}, D_{i+1}, ..., D_n\}$ and cost functions $\{C_j\}_{j=1}^n$ such that $D_i = \max_j\{D_j\}$ and

$$\sum_{j=1}^{n} C_j(q_j) \geq \epsilon + \Big(1 + \frac{d}{-d + \sum_{j \neq i} D_j}\Big) \sum_{j=1}^{n} C_j(q_j^*).$$

Theorem Implication - First, note that the result, i.e., the bounds, in the theorem depend only on $\{D_i\}_{i=1}^n$, and d, and is independent of $\{C_i\}_{i=1}^n$. The theorem helps us derive insights on the effect of the size of market participants on the efficiency loss. Given a fixed aggregate privacy compromise demand d, we observe that the efficiency loss upper bound is increasing with the capacity limit of the DH with the largest capacity D_i (due to incurring more cost on compromising more), and is decreasing with the total capacity, $\sum_{j \neq i} D_j$ of the other DHs (due to implication of Corollary 5). In the case when $\sum_{j \neq i} D_j - d$ approaches zero (following Lemma 5), it is possible that an ONE leads to high efficiency loss. However, our proposed market mechanism *guarantees approximate social optimality* at all ONE, if there *exist at least two DHs with large compromise capacity limits, or a large number of DHs with small compromise limits (following Lemmas 4 and 5).* As an example if there are m where $m \leq n$ DHs each with a compromise limit larger than

[5]According to this assumption, the marginal cost of every DH i at a zero compromise level, $\partial^- C_i(0)$, should be no greater than the marginal cost of a DH who compromises a positive amount at the social optimum. This will always hold true if $\partial^- C_i(0) \leq \partial^+ C_j(0)$, for every pair of DHs i and j. Here, ∂^- and ∂^+ denote the left and right directional derivatives of C_i as the latter might not be differentiable.

$\epsilon d > 0$, the ratio of the aggregate cost at ONE, and that resulting from a socially optimal state is upper bounded by

$$1 + \frac{1}{\epsilon(m-1) - 1},$$

that converges to 1 as m grows large; this can be deduced directly from Theorem 8.

Measuring Market Power of DHs - In view of the market efficiency, we consider determining the *Lerner's index* [9], which is the indicator market power of DH i, and is defined below:

$$LR_i(b) = \frac{p(b) - \frac{\partial^+ C_i(S_i(b_i, p(b)))}{\partial q_i}}{p(b)},$$

where $p(b)$ is the market clearing benefit resulting from bid vector b, and $\partial^+ C_i(S_i(b_i, p(b)))$ is the right directional derivative of DH i's cost at her privacy compromise level $(S_i(b_i, p(b)))$. LR_i lies in the range $[0, 1]$, with a higher Lerner index being an indicator of higher market power. We have the following result related to bounds of DH Lerner index. the proof of which is in the Appendix.

Corollary 6. *Let C_i be continuous and convex, and strictly increasing in $[0, \infty)$ with $C_i(0) = 0$, $\sum_{j \neq i} D_j > d$, for every DH i. At an ONE b, if DH i's privacy compromise amount is less than D_i, we have*

$$LR_i(b) \leq \frac{D_i}{-d + \sum_j D_j}$$

Corollary Implication - We see that even when DHs have a high cost and provide a low compromise amount, the market-clearing benefit cannot be too high due to every DH's Lerner index being bounded. We have already seen that at an ONE, at least two DH's compromise amounts are less than their capacity limits. Thus, the bound derived from the corollary is applicable to at least two DHs. Let DH i be one such DH whose compromise supply is less than its capacity threshold. In the event that the capacity threshold of all DHs grows much larger than d (something of interest to policy and incentive makers), the Lerner index bound in the corollary converges to zero. This implies that the marginal cost of DH i approximately equals the market-clearing benefit, and every DH has nearly equal market power.

5 Distributed Bidding Algorithms

In the previous sections, we focused on analyzing privacy trading markets for market equilibria when DHs have (and do not have) upper and lower bounds on their privacy compromise amounts. In this section, our focus is to develop supply bidding algorithms that converge to market equilibria for perfectly competitive and oligopolistic markets that function in a distributed manner, and scale well with the number of DHs. 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, thereby facilitating scaling, as mentioned above. As potential candidate algorithm types, one could either use the standard dual gradient algorithm proposed in [20], or the alternative direction multiplier method in [21]. Both types are iterative in nature, and equivalently maps the supply bidding process. In this work, we resort to the dual gradient algorithm in [20], without loss of generality. *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. Consequently, our proposed distributed bidding algorithms possess all the convergence properties of dual gradient algorithm. We refer the readers to [20] for details regarding optimal step sizes, the stopping criterion, and convergence speed. As an example of the high convergence speed, we show in the following section that for very low γ values in Algorithms 1 and 2, convergence is very fast even for a large number of DHs. To be more specific, it is shown in [20] that very small γ values result in an exponential convergence rate.

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

$$b_i(k) = \left[\frac{(C_i')^{-1}(p(k))}{p(k)}\right]^+ \tag{16}$$

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

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

$$p(k+1) = \left[p(k) - r\left(\sum_i b_i(k)p(k) - d\right)\right]^+ \tag{17}$$

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 [20] and, repeat

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

$$b_i(k) = \left[\frac{(D_i')^{-1}(p(k))}{p(k)}\right]^+ \tag{18}$$

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

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

$$p(k+1) = \left[p(k) - r\left(\sum_i b_i(k)p(k) - d\right)\right]^+ \tag{19}$$

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 [20] and, repeat

6 Numerical Evaluation

In this section, we run numerical experiments to study the following aspects of market behavior not covered through theoretical results in the paper: (a) the speed of convergence

of our proposed supply bidding algorithms, and their scalability with increasing number of DHs, (b) the effect of DH cost functions on the amount of efficiency loss in the ONE, and (c) for the specific case when privacy compromise limits exist, the study of market equilibrium parameters and social welfare optimality under conditions varying in the number of DHs that have small compromise limits, amidst few DHs having relatively large compromise limits.

6.1 Evaluation Setup

To study (a) above, 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 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[6] [12] we define to be $\frac{MI(X_i;Y_i)}{H(X_i)}$, where X_i is the source distribution[7] 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 [22] of X_i, and $MI(X_i;Y_i)$ is the mutual information [22] 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 others have their a_i and h_i coefficients set high and randomly selected in the interval $[1, 2]$. In order to study (c), we vary the number of small (in terms of compromise limits) DHs between two and 10 and fix the number of relatively large DHs to two. The small DHs have a compromise limit of one privacy leakage unit while the relatively large DHs have capacity limits of 10 privacy leakage units.

6.2 Evaluation Analysis

We use the standard and widely popular *tâtonnement process* [23][24] to converge to market equilibrium in a computational manner. We observe from Figures[8] 2a and 2b (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

[6] Our methodology is general and independent of the information-theoretic privacy metric.

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

[8] Each figure is a representative of 50 instances of a numerical experiment.

RAM) to the market equilibrium (PCE and ONE respectively). 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 2c and 2d (where $\gamma = 0.05$), we show the scalability of Algorithms 1 and 2. The results and rationale are very similar to those in Figures 2a and 2b, and convergence to market equilibrium is equally fast.

Figure 2e plots the comparison of benefit and total cost respectively at PCE and ONE. Figure 2f plots the amount of privacy compromise by low and high cost users respectively, at PCE and ONE. We observe form Figure 2e that if all DHs are homogeneous, 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 mentioned in the evaluation setup, we observe from Figure 2f that the differences between market equilibria decrease quickly with increase in the number of DHs. This is due to the fact that with increase in market size, 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 between the two market equilibria are large when one DH has a low cost function - this is because the latter has market power. 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 more cost competing DHs into the market to improve social welfare. When the market size is large, the differences between the two market equilibria are small in all the three cases. 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 with low 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 3. 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 2.

We study via Figure 4, the case when DHs have privacy compromise limits. More specifically we investigate market equilibrium parameters and social welfare optimality under conditions varying in the number of DHs that have small compromise limits, amidst few DHs having relatively large compromise limits. We first consider the case *when the largest DH has the lowest marginal cost* by considering $C_1(q)$ to be linear in q with unit slope. The first thing to note that needs no experimentation is that the Lerner's index is the same for all small DHs of the same compromise limit as they have the same marginal cost. Through experiments we observe that as the number of small DHs increase, the

largest DH provides more compromise, as a result the equilibrium allocation becomes more efficient with increasing small DHs. There is an exception though - the supply provided when the number of small DHs equals 4, because for both cases of the number of small DHs being 2 and 4 respectively, the compromise supply provided by these small DHs reach capacity limits. We also observe that the market benefit and the Lerner indices decrease with the size of small DHs, due to increased market competition. In addition, with higher values of D_1, the market benefit decreases; the intuition following directly from (12). However, a higher value of D_1 always leads to higher aggregate cost at ONE. This is because a higher value of D_1 always results in less compromise supply from DH 1 (follows from Corollary 5). Finally, the numerical experiments verify Theorem 8, i.e., the ratio of the social welfare at ONE to the optimal social welfare is upper bounded as mentioned in Theorem 8.

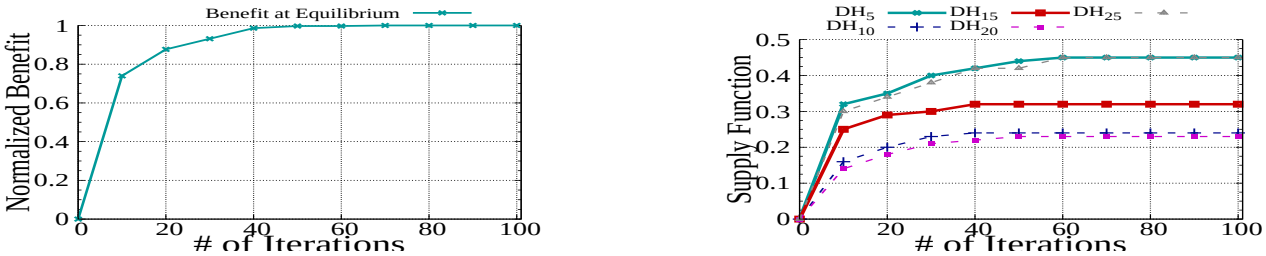
We now consider the case *when the largest DH has the highest marginal cost* by considering $C_1(q)$ to be linear in q with a slope (without loss of generality) of 2.5, all other parameters remaining the same as with the case when the largest DH has the lowest marginal cost. It is obvious that the social welfare is maximized when the largest DH provides no compromise. Thus the efficiency loss at ONE with $C_1(q) = 2.5q$ is much smaller than that with $C_1(q) = q$. We also observe that the market benefit and the Lerner indices decrease with the number of DHs. However, unlike the case mentioned in the aforementioned paragraph, an increase in D_1 increases the social welfare when the number of small DHs equals 2. This is because a higher value of D_1 reduces the supply from DH_1 (follows from Corollary 5). It also leads to lower market benefits and lower Lerner indices.

7 Related Literature

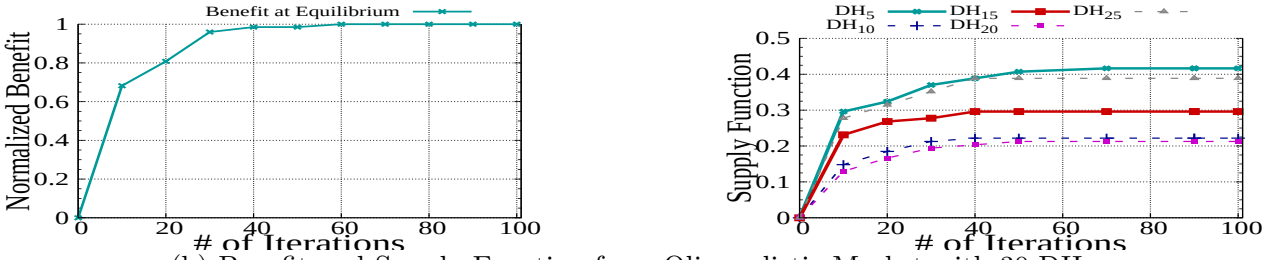
Research on privacy trading markets is scarce, being a fairly recent topic. In this section, we briefly review related literature most relevant to privacy trading markets - primarily covering the area of differential privacy and mechanism design, and propose differences in our approach inline where applicable. We emphasize here that our work is applicable to general information gain privacy metrics as mentioned in [12], including differential privacy.

Most existing works on privacy-aware mechanism design [25][26][27][28][29][30][31][32] assume that there is a trusted data holder. The private data is either already kept by the data holder, 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, and may release it to agencies like ad-networks in return for benefits.* To this end, in the seminal work by [26], 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 and is not for release to an ad-network, and (b) cost functions need not be linear but convex.*

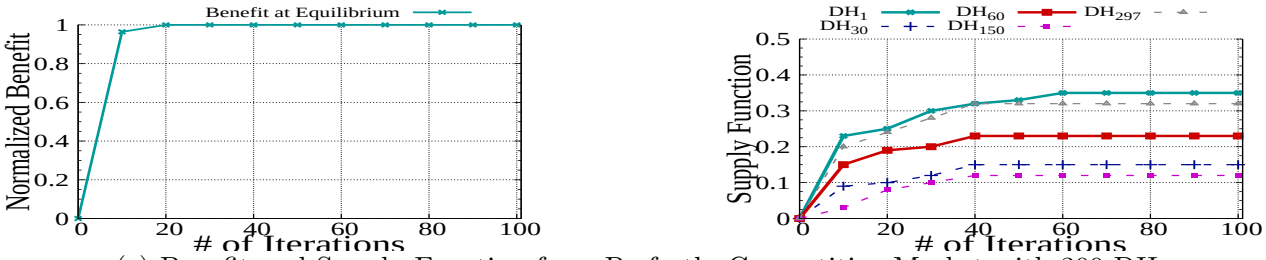
Subsequent works [27][28][29][31] explore various models for individuals' (analogous to DHs in our work) valuation of privacy, especially the correlation between the cost functions and the private bits. This line of work has been extended to the scenario that the data is



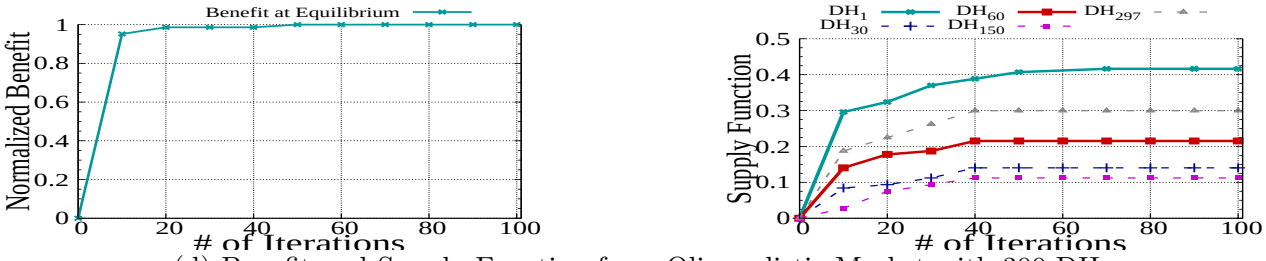
(a) Benefit and Supply Function for a Perfectly Competitive Market with 30 DHs



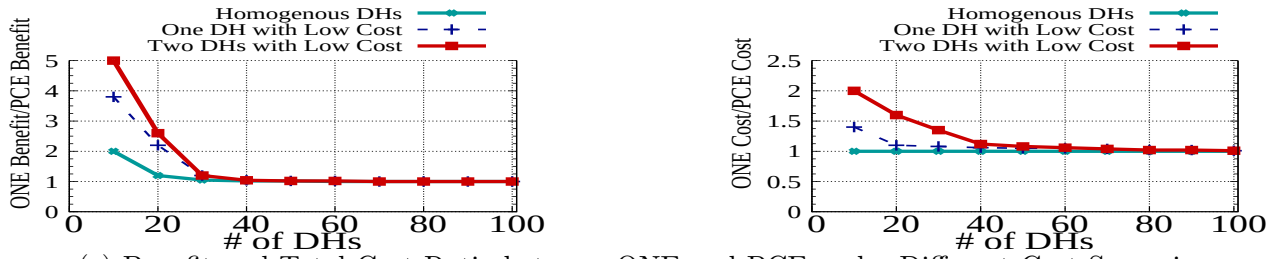
(b) Benefit and Supply Function for a Oligopolistic Market with 30 DHs



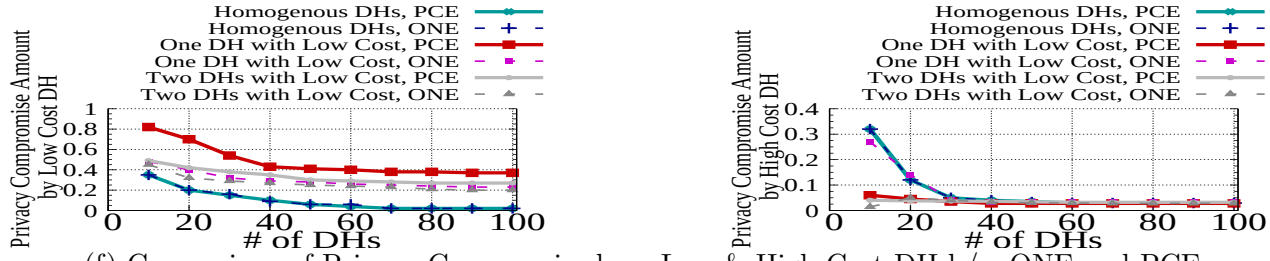
(c) Benefit and Supply Function for a Perfectly Competitive Market with 300 DHs



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

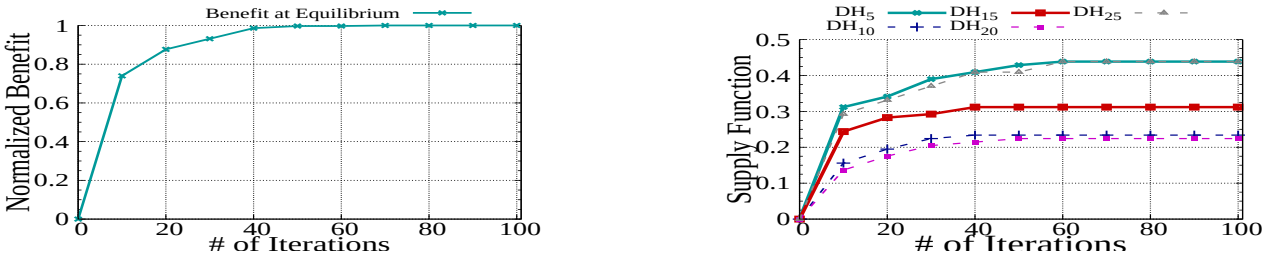


(e) Benefit and Total Cost Ratio between ONE and PCE under Different Cost Scenarios

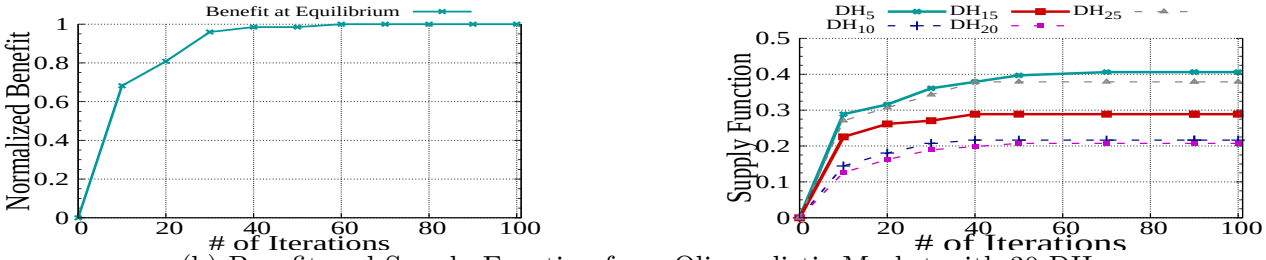


(f) Comparison of Privacy Compromise by a Low & High Cost DH b/w ONE and PCE

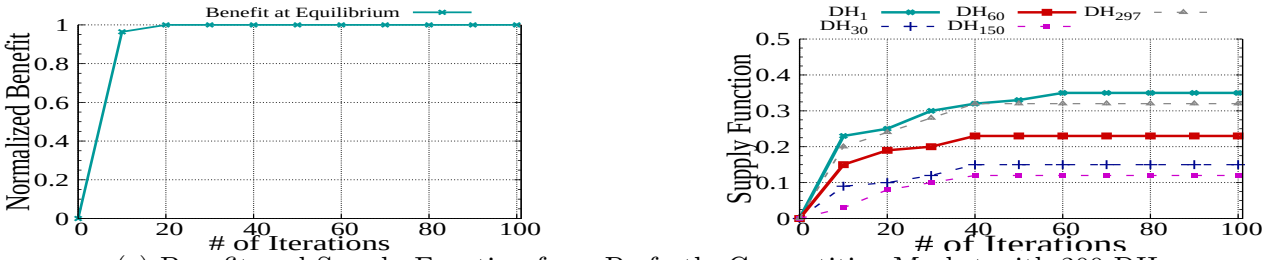
Figure 2: Comparison of Market Properties with ***Linear Marginal DH Cost***



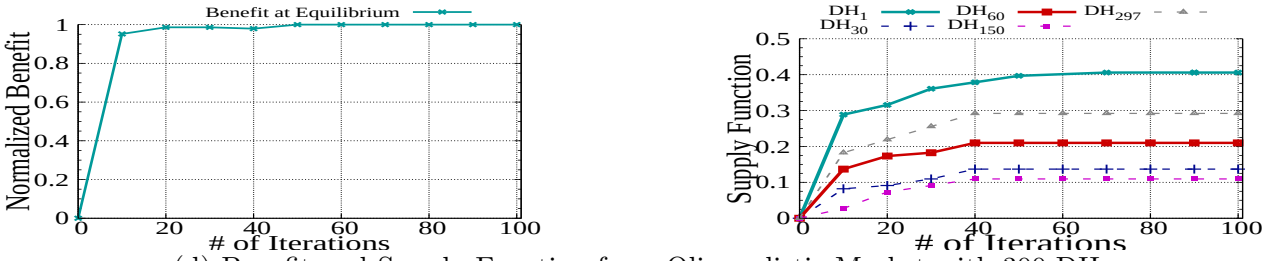
(a) Benefit and Supply Function for a Perfectly Competitive Market with 30 DHs



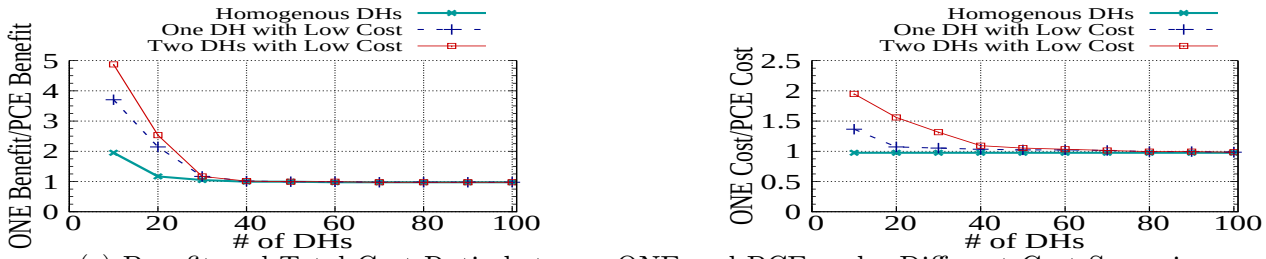
(b) Benefit and Supply Function for a Oligopolistic Market with 30 DHs



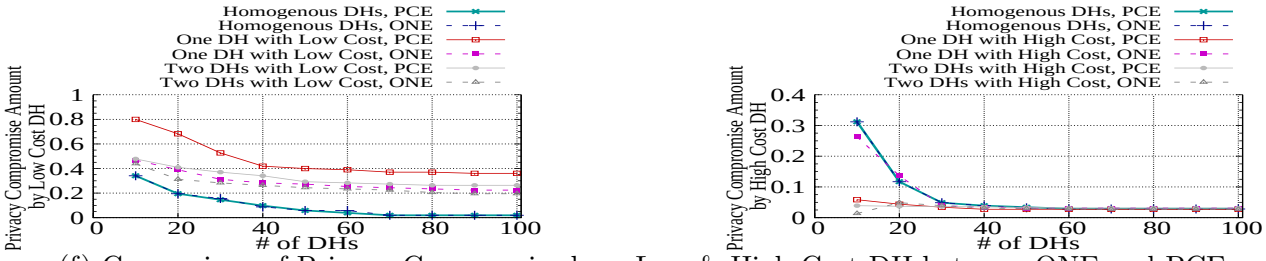
(c) Benefit and Supply Function for a Perfectly Competitive Market with 300 DHs



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

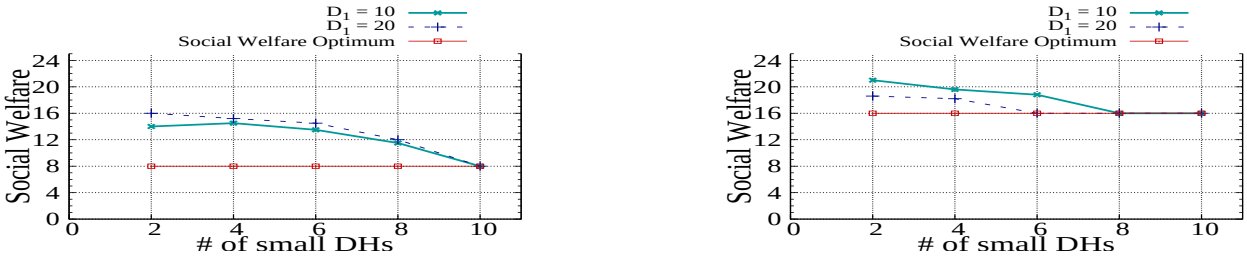


(e) Benefit and Total Cost Ratio between ONE and PCE under Different Cost Scenarios

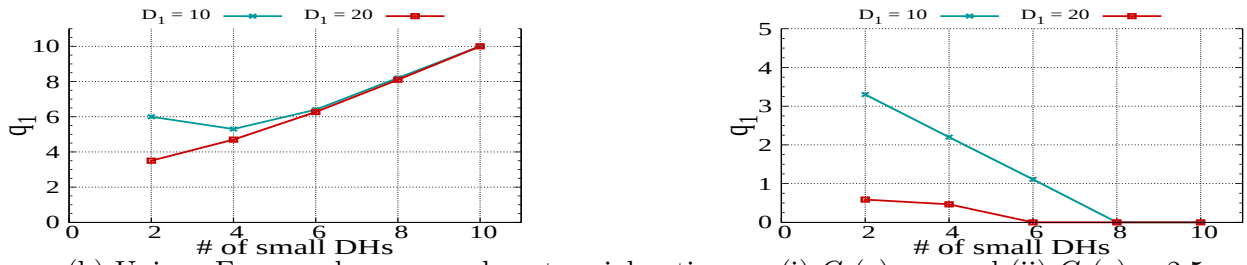


(f) Comparison of Privacy Compromise by a Low & High Cost DH between ONE and PCE

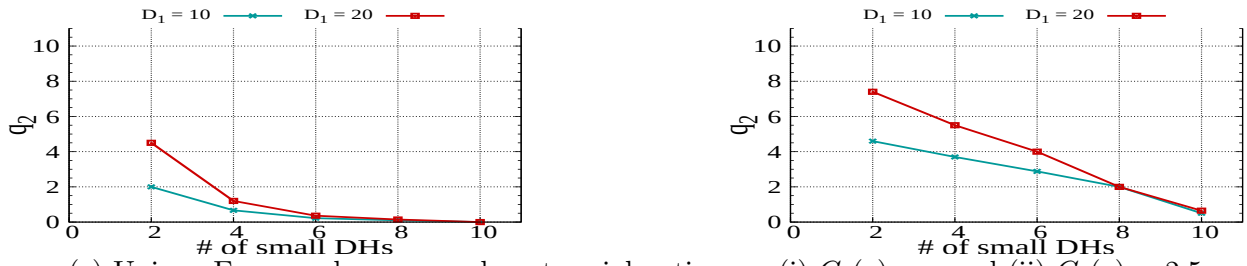
Figure 3: Comparison of Market Properties with ***Constant Marginal DH Cost***



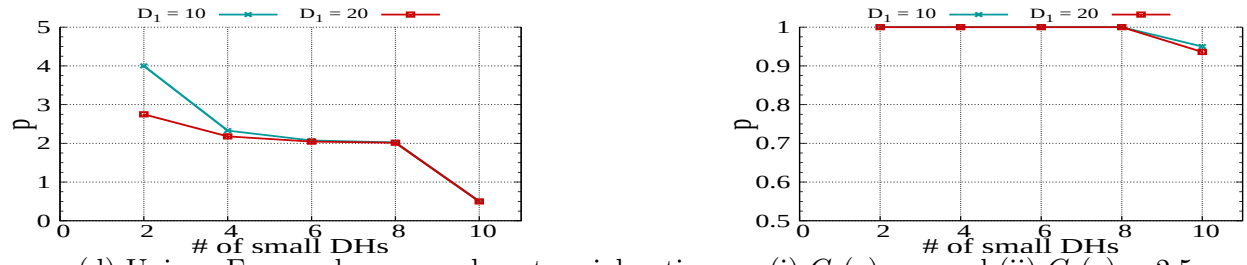
(a) Total Aggregate Cost at the Unique ONE vs Socially Optimal Total Cost: (i) $C_1(q) = q$ and (ii) $C_1(q) = 2.5q$



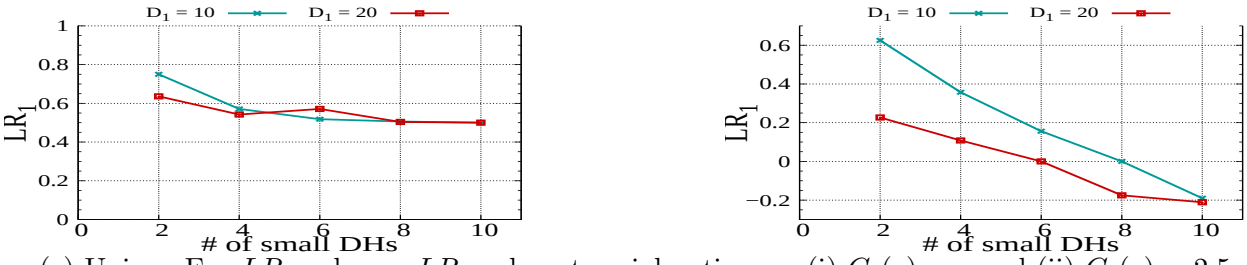
(b) Unique Eq. q_1 value vs q_1 value at social optimum: (i) $C_1(q) = q$ and (ii) $C_1(q) = 2.5q$



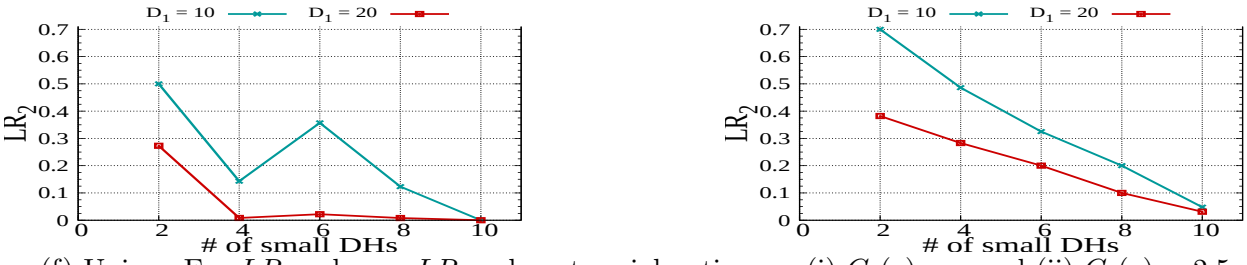
(c) Unique Eq. q_2 value vs q_2 value at social optimum: (i) $C_1(q) = q$ and (ii) $C_1(q) = 2.5q$



(d) Unique Eq. p value vs p value at social optimum: (i) $C_1(q) = q$ and (ii) $C_1(q) = 2.5q$



(e) Unique Eq. LR_1 value vs LR_1 value at social optimum: (i) $C_1(q) = q$ and (ii) $C_1(q) = 2.5q$



(f) Unique Eq. LR_2 value vs LR_2 value at social optimum: (i) $C_1(q) = q$ and (ii) $C_1(q) = 2.5q$

Figure 4: Comparison of Market Properties at ONE vs at the Social Optimal Point: (i) $C_1(q) = q$ and (ii) $C_1(q) = 2.5q$

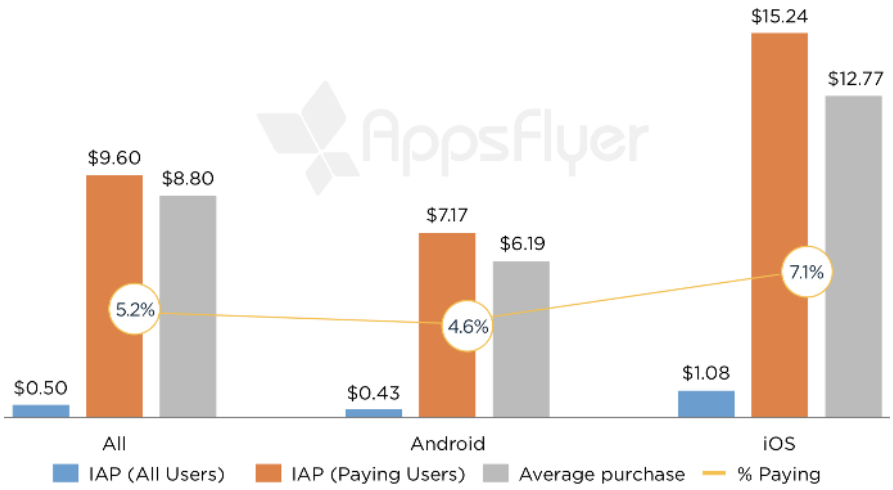


Figure 5: The State of In-App Spending, 2016, Source: AppsFlyer

not available yet and needs to be reported by the individuals to the data collector, but the data collector is still trusted [30][33][34][32] - *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, [13] gives a comprehensive survey. In [35], 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.* As an example of a practical realization of a privacy market, a recent project: HAT Data Exchange (HATDex), by Ng, 2018 [45] advocates the ecosystem of person-controlled personal data (PPD) instead of organization-controlled personal data (OPD).

8 Discussion

We proposed a privacy trade market framework, *Privacy Bazaar*, for mobile apps and IoT ecosystems that aims to maximize utilitarian social welfare amongst competing data holders (e.g., apps, IoT boxes, PDSs) by preserving their *heterogeneous* privacy preservation constraints upto certain compromise levels, induced by their clients, and at the same time satisfying requirements of ad-networks. *One could argue against the design of the mechanism proposed in the paper, and vouch for the case of paid apps to give users an ad-free environment or an environment of viewing significantly less ads.* To this end, in Section 8.1, we provide rationale behind backing the necessity of our model, backed by real statistical data. In addition, we emphasize that despite the important role DHs can and should play in protecting user's privacy (with or without privacy trading), we can not put complete faith in them for user's privacy protection in the long term. This assertion of ours is confirmed by recent events like the Facebook-Cambridge Analytica data scandal where personal information of 87 million users were shared with third parties without users' knowledge [2], Google providing third parties access to users' Gmail

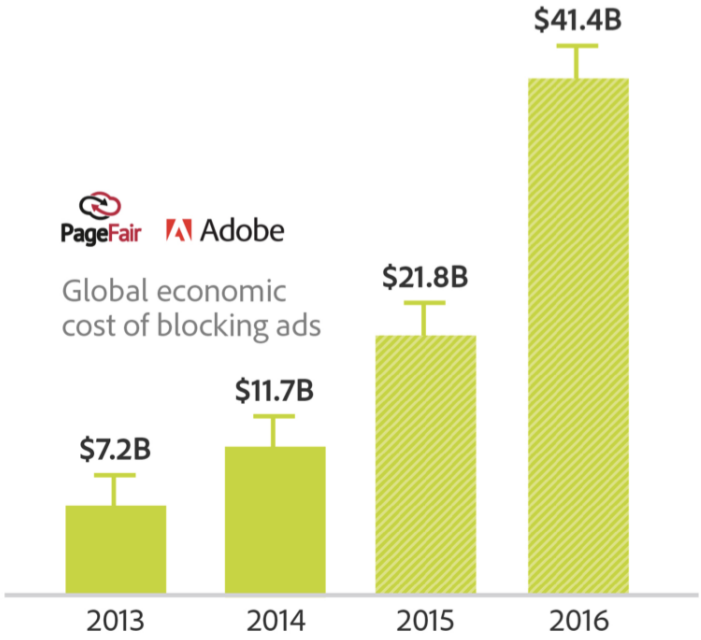


Figure 6: The Cost of Blocking Ads, Source: pagefair.com

account without users' explicit consent [36]. As a result, one could vouch for end users needing to take control of their privacy (e.g., via user consent in data release as mandated by the recently operative GDPR). Subsequently, Section 8.2 discusses the challenges in realizing this objective, irrespective of the presence and absence of GDPR like regimes.

8.1 Paid Apps Versus Unpaid Apps

In the free mobile apps ecosystem, apps collect user's information in order to generate revenue by trading these information with third parties like ad-networks. This increases the risk of end-user privacy breaches. Another alternative for apps to generate revenue is for them to charge some nominal amount from end-users instead of trading their personal information with third parties. However, a recent study by *AppsFlyer* [37] suggests that although users may care about their privacy, many of them are not willing to pay money for using apps. This situation gets only exacerbated considering the recent finding that Android users have an average of 95 apps installed on their phones, 35 of which are used (on average) each day [38]. According to *AppsFlyer* study, the average monthly in-app purchase per user globally is $1.08 for iOS users and $0.43 for Android users (see Figure 5). In Asia, the average per user in-app purchase amounts to $1.74 for iOS users and $0.44 for Android users. In North America, the same purchase amounts to $0.79 for iOS users and $0.46 for Android users. In Europe, the average in-app purchase per user amounts to $0.36 for iOS users and $0.23 for Android users. Finally, in Latin America the same purchase amounts to $0.32 for iOS users and $0.09 for Android users. Data is not available for the Australian continent. [37][39].

These above-mentioned statistics suggest that it is unrealistic to expect significant changes at the users' end in near future with respect to app payment behavior. Therefore, changes in ecosystem should be made at the end of other players like app-developers (or data holders), advertisers and ad-network, so as to respect user privacy. As an example, it is easier to enforce such changes at these ends through regulations [40][41].

A recent article in *The Economist* [44] has suggested apps paying users for their data (aligned with our methodology in this paper, and something that can be envisioned

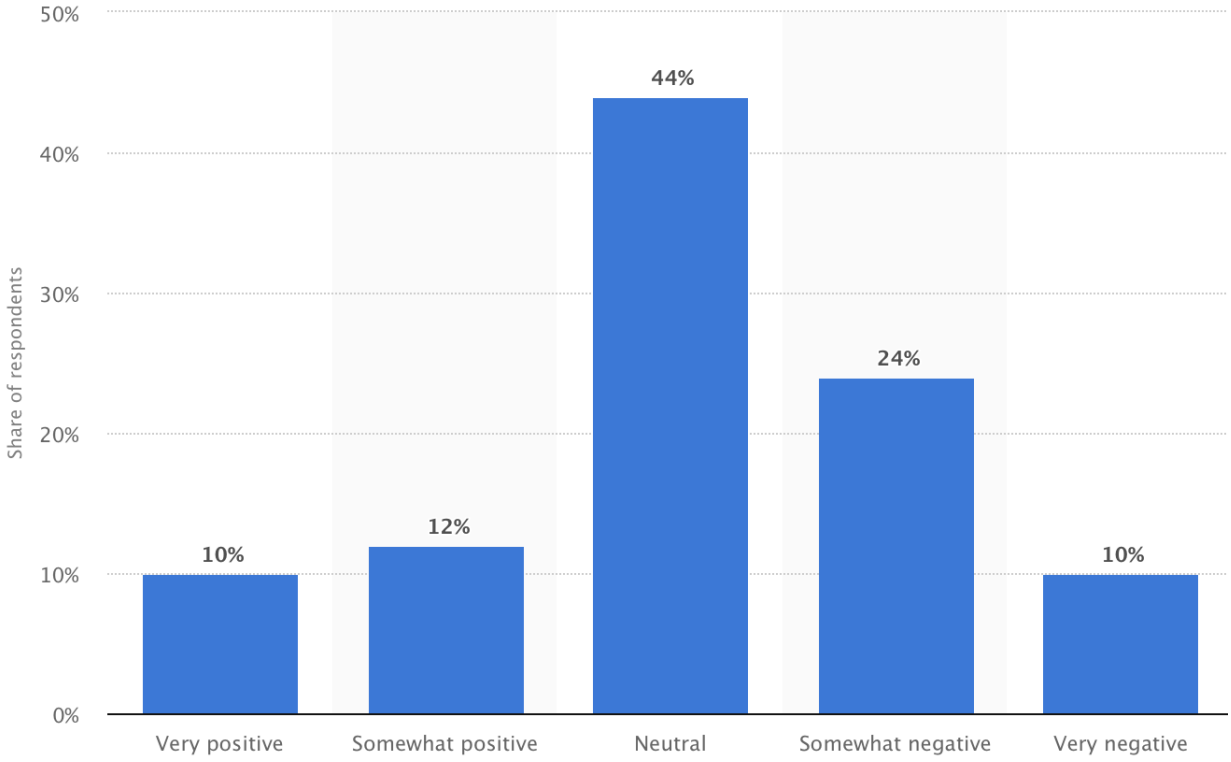


Figure 7: Attitude of Americans towards websites collecting cookies, Source: Statistic.com

to be implemented under the GDPR via user consent), when keeping apps free of cost (assuming that paid apps does not result in ad viewing). However, to achieve this goal, users first need to have full control over their data and its flow, which is missing in the current ecosystem (and potentially infeasible), else apps can get user data from other sources for free or at a cheaper rate. This idea is closely related to another recent project: HATDex [45] that advocates the culture of person-controlled personal (PPD) data instead of organization-controlled personal data (OPD). However, the authors argue that due to the inability of users having full control of their data and its flow, incomplete data trading contracts will result and subsequently social welfare at market equilibrium will not be optimized due to externalities not being completely internalized. In the most optimistic setting, even if users get money for use of their data, statistics show that they would not get much, e.g., a typical Facebook user would get only $9 per year [44]. So, instead of paying users for their data, providing other non-monetary benefits (also aligned with our methodology)like personalized services may be a better alternative.

8.2 Challenges to End-User Controlling Privacy

In non-GDPR settings, where user consent[9] for getting its data may not be mandatory, we feel that steps should be taken to give more privacy control to end-users, despite them not willing to pay for apps. Recent works in field of usable security and privacy have focused on end-users to achieve better privacy [46]. They aim to nudge/teach users in making choices which are safe from privacy perspective. As an example, most of current free apps request users for a number of permissions in their generated default list that

[9]According to renowned philosopher Onere O'Neill, providing user consent to applications is a function of the interactions one has with its social environment and the trust one puts on social colleagues. Thus, being influenced by non-judicious people might result in giving consent to wrong set of applications.

are not necessarily required for app functioning, and users tend to approve the default list (often due to a binary all or none 'agree - disagree' clause to force needy users of a service to press the 'accept' button) of requested permissions during installation period [47]. Later on these extra permissions are exploited to collect private information. Nudge theory [48] attempts to make users aware of these extra non-required permissions [49]. It has produced desired results in short-term. However, it remains an open challenge to achieve similar results in the long-term since users are quite forgetful, are prone to habituation, and make irrational choices [50][51]. Another important challenge to giving user the power to control privacy is the well known privacy paradox [52][53]. According to this paradox, people express genuine concerns about their online privacy, yet continue to broadcast personal details in public forums and on websites that warn them that they are collecting their data [52]. Figure 7 showing 66 % respondents having non-negative attitude towards websites collecting cookies puts weight to this assertion.

In recent years, ad blockers have become quite popular. From end-users' perspective, ad blocker is an powerful tool which can be used to limit the number of ads to be shown on their devices. According to a report from *PageFair* [42][43], ad blocker usage surged 30% in 2016. There were 615 million devices blocking ads worldwide by the end of 2016 and 62% (308 million) of those were mobile. In 2016, there were 4.3 billion mobile phone users[10]. So, it can be safely inferred that penetration of ad blockers among mobile phone users is not significantly high. However, the increased usage of ad blockers has indeed affected the revenue of players involved in the ad business. Due to increased usage of ad blockers, players in ad business have lost revenue worth billions of dollars (see Figure 6). This solution (of using ad blockers) may be appropriate for end users from a privacy perspective but it is not sustainable for the free mobile-app ecosystem and all players involved in it, in the long run. We believe that any new solution must consider the interests of all players, as we have proposed, rather than considering interests of just one (or few) player.

9 Summary and Future Work

In this paper, we proposed *Privacy Bazaar*, a rigorous privacy trade market model for mobile apps and IoT ecosystems that aims to achieve a maximum social welfare state amongst competing data holders (e.g., apps, IoT boxes, PDSs) 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. To this end, using concepts from parameterized supply-function economics, we proposed the first mathematically rigorous privacy market design paradigm 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. Consequently, we characterized the efficiency gap in closed form. We also proposed scalable distributed supply function bidding algorithms that converge to market equilibria exponentially fast.

As part of future work, we are interested to study the impact of schemes like GDPR on the privacy trading business, e.g., in terms of revenue loss for advertisers and ad-networks. In addition, we wish to analyze the role of multiple competing ad-networks, a network of ad-networks, and one-many consumer-DH mappings on privacy trade. Regu-

[10]https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/

latory agencies play a vital role in enforcing privacy measures. We want to understand implications of regulatory measures enforced by such agencies. A vital research direction that needs attention is the design of mechanisms that close the efficiency gap in oligopolistic privacy trading markets. In this regard, we aim to explore the design of behavioral economic mechanisms that achieve this goal. We want to address the issue of fairness in privacy trading, whereby we want to characterize the deviation of a social welfare optimal state from a state where the privacy compromise process is done in a mathematically fair manner among DHs. This study will have implications in the design of privacy trading mechanisms in capitalist as well as socialist economies. Finally, we want to explore the role and impact of privacy enhancing technologies in the formation and evolution of privacy markets. As a future research topic, the sharing and aggregation of machine learning models based on the raw data and other models is a very important direction.

References

[1] Meng, W., Ding, R., Chung, S. P., Han, S., & Lee, W. (2016, February). The Price of Free: Privacy Leakage in Personalized Mobile In-Apps Ads. In NDSS.

[2] Wikipedia (2018) Facebook-Cambridge Analytica data scandal, [Online accessed: 19-September-2018].

[3] Mortier, R. M., Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Glover, K., & Wang, L. (2018). Building Accountability into the Internet of Things: The IoT Databox Model.

[4] The Economist (2018) Google and Mastercard cut a secret ad deal to track retail sales, [Online accessed: 19-September-2018].

[5] Claburn, T., (2017) Ad blocking basically doesn't exist on mobile, In The Register.

[6] Ikram, M., & Kaafar, M. A. (2017). A First Look at Ad Blocking Apps on Google Play. arXiv preprint arXiv:1709.02901.

[7] Perera, C., Wakenshaw, S. Y., Baarslag, T., Haddadi, H., Bandara, A. K., Mortier, R., & Crowcroft, J. (2017). Valorising the IoT databox: creating value for everyone. Transactions on Emerging Telecommunications Technologies, 28(1), e3125.

[8] Klemperer, P. D., & Meyer, M. A. (1989). Supply function equilibria in oligopoly under uncertainty. Econometrica: Journal of the Econometric Society, 1243-1277.

[9] Johari, R., & Tsitsiklis, J. N. (2011). Parameterized supply function bidding: Equilibrium and efficiency. Operations research, 59(5), 1079-1089.

[10] Guardian News Editorial (2018). The Guardian View on Internet Privacy: It's the Psychology Stupid.

[11] Roth, A. E., (2015). Who Gets What?and Why: The New Economics of Matchmaking and Market Design, Houghton Mifflin Harcourt.

[12] Wagner, I., & Eckhoff, D. (2018). Technical privacy metrics: a systematic survey. ACM Computing Surveys (CSUR), 51(3), 57.

[13] Pai, M. M., & Roth, A. (2013). Privacy and mechanism design. ACM SIGecom Exchanges, 12(1), 8-29.

[14] Hobbs, B. F., Rothkopf, M. H., Hyde, L. C., & O'neill, R. P. (2000). Evaluation of a truthful revelation auction in the context of energy markets with nonconcave benefits. Journal of Regulatory Economics, 18(1), 5-32.

[15] Ausubel, L. M., & Milgrom, P. (2006). The lovely but lonely Vickrey auction. Combinatorial auctions, 17, 22-26.

[16] Rothkopf, M. H., Teisberg, T. J., & Kahn, E. P. (1990). Why are Vickrey auctions rare?. Journal of Political Economy, 98(1), 94-109.

[17] Maheswaran, R. T., & Basar, T. (2004, December). Social welfare of selfish agents: motivating efficiency for divisible resources. In Proceedings of 43rd IEEE Conference on Decision and Control.

[18] Johari, R., & Tsitsiklis, J. N. (2009). Efficiency of scalar-parameterized mechanisms. Operations Research, 57(4), 823-839.

[19] Roughgarden, T. (2005). Selfish routing and the price of anarchy (Vol. 174). Cambridge: MIT press.

[20] Bertsekas, D. P., & Tsitsiklis, J. N. (1989). Parallel and distributed computation: numerical methods (Vol. 23). Englewood Cliffs, NJ: Prentice hall.

[21] Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine learning, 3(1), 1-122.

[22] M. Cover and J. A. Thomas, Elements of information theory. John Wiley & Sons, 2012.

[23] Varian, H. R. (1992). Microeconomic analysis.

[24] Arrow, K., & Intriligator, M. (2000). Handbook of mathematical economics (Vol. 1). Elsevier.

[25] Pal, R., Hui, P., & Prasanna, V. K. (2018). On Optimal Privacy Engineering for the Smart Micro-Grid. IEEE Transactions on Knowledge and Data Engineering.

[26] Ghosh, A., & Roth, A. (2015). Selling privacy at auction. Games and Economic Behavior, 91, 334-346.

[27] Fleischer, L. K., & Lyu, Y. H. (2012, June). Approximately optimal auctions for selling privacy when costs are correlated with data. In Proceedings of the 13th ACM Conference on Electronic Commerce (pp. 568-585). ACM.

[28] Ligett, K., & Roth, A. (2012, December). Take it or leave it: Running a survey when privacy comes at a cost. In International Workshop on Internet and Network Economics (pp. 378-391). Springer, Berlin, Heidelberg.

[29] Roth, A., & Schoenebeck, G. (2012, June). Conducting truthful surveys, cheaply. In Proceedings of the 13th ACM Conference on Electronic Commerce (pp. 826-843). ACM.

[30] Ghosh, A., & Ligett, K. (2013, June). Privacy and coordination: computing on databases with endogenous participation. In Proceedings of the fourteenth ACM conference on Electronic commerce (pp. 543-560). ACM.

[31] Nissim, K., Vadhan, S., & Xiao, D. (2014, January). Redrawing the boundaries on purchasing data from privacy-sensitive individuals. In Proceedings of the 5th conference on Innovations in theoretical computer science (pp. 411-422). ACM.

[32] Ghosh, A., Ligett, K., Roth, A., & Schoenebeck, G. (2014, June). Buying private data without verification. In Proceedings of the fifteenth ACM conference on Economics and computation (pp. 931-948). ACM.

[33] Xiao, D. (2013, January). Is privacy compatible with truthfulness?. In Proceedings of the 4th conference on Innovations in Theoretical Computer Science (pp. 67-86). ACM.

[34] Chen, Y., Chong, S., Kash, I. A., Moran, T., & Vadhan, S. (2016). Truthful mechanisms for agents that value privacy. ACM Transactions on Economics and Computation (TEAC), 4(3), 13.

[35] Wang, W., Ying, L., & Zhang, J. (2016, June). The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits. In ACM SIGMETRICS Performance Evaluation Review (Vol. 44, No. 1, pp. 249-260). ACM.

[36] The Independent UK (2018) Google admits giving hundreds of firms access to your gmail account, `https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-gmail-data-sharing-email-inbox-privacy-scandal-a8548941.html`, [Online accessed: 23-September-2018].

[37] AppsFlyer (2016) The State of In-App Spending: 2016, Benchmarks Regional & Global, [Online accessed: 19-September-2018].

[38] The Next Web (2014) Android users have an average of 95 apps installed on their phones, according to Yahoo Aviate data, [Online accessed: 23-September-2018].

[39] Mary Kearl (2017) 30 Essential Stats On In-App Purchases And Monetization, In Braze Magazine[Online accessed: 19-September-2018].

[40] Posner, R. A. (1974). Theories of economic regulation.

[41] Milberg, S. J., Smith, H. J., & Burke, S. J. (2000). Information privacy: Corporate management and national regulation. Organization science, 11(1), 35-57.

[42] FairPage (2016) The cost of ad blocking, [Online accessed: 23-September-2018].

[43] Business Insider UK (2016) Ad blocker usage is up 30% and a popular method publishers use to thwart it isn't working, [Online accessed: 23-September-2018].

[44] The Economist (2018). What if people were paid for their data, 11(1), [Online accessed: 19-September-2018].

[45] Ng, I. C. (2018). The Market for Person-controlled Personal Data with the Hub-of-all-Things (HAT), Working Paper, Warwick Manufacturing Group. WMG Service Systems Research Group Working Paper Series (01/18), Unpublished.

[46] Liu, B., Andersen, M. S., Schaub, F., Almuhimedi, H., Zhang, S. A., Sadeh, N., & Acquisti, A. (2016, June). Follow my recommendations: A personalized privacy assistant for mobile app permissions. In Symposium on Usable Privacy and Security.

[47] Leontiadis, I., Efstratiou, C., Picone, M., & Mascolo, C. (2012, February). Don't kill my ads!: balancing privacy in an ad-supported mobile application market. In Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications (p. 2). ACM.

[48] Leonard, T. C. (2008). Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness.

[49] Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., & Agarwal, Y. (2015, April). Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In Proceedings of the 33rd annual ACM conference on human factors in computing systems (pp. 787-796). ACM.

[50] Kahneman, D., & Egan, P. (2011). Thinking, fast and slow (Vol. 1). New York: Farrar, Straus and Giroux.

[51] Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. Psychological review, 73(1), 16.

[52] Barnes, Susan B. "A privacy paradox: Social networking in the United States." First Monday 11.9 (2006).

[53] Hallam, Cory, and Gianluca Zanella. "Wearable device data and privacy: A study of perception and behavior." World 7.1 (2016).

[54] Nash, J. (1951). Non-cooperative games. Annals of mathematics, 286-295.

[55] Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). Microeconomic theory (Vol. 1). New York: Oxford university press.

[56] Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). Microeconomic theory (Vol. 1). New York: Oxford university press.

[57] Varian, H. R., & Harris, C. (2014). The VCG auction in theory and practice. American Economic Review, 104(5), 442-45.

[58] K. J. Arrow, Social choice and individual values, vol. 12. Yale university press, 2012.

[59] Myerson, R. B. (1981). Utilitarianism, egalitarianism, and the timing effect in social choice problems. Econometrica: Journal of the Econometric Society, 883-897.

A Glossary of Definitions

In this section, we provide brief description of some economic terminologies used in the paper.

- **Nash Equilibrium** is a stable state of a system involving the interaction of different participants, in which no participant can gain by a unilateral change of strategy if the strategies of the others remain unchanged [54]. In our work, we use this concept to characterize a stable state in which no data holder has incentive to make changes in its privacy compromise amount unilaterally.

- **Perfect Competition** is a state prevailing in a market in which buyers and sellers are so numerous and well informed that all elements of monopoly are absent and individual buyers and sellers can not influence the market price [55]. In our work, we show that, in a perfectly competitive market, our proposed mechanism can converge to a unique equilibrium which maximizes welfare of all players involved. We consider this market to realize ideal benchmarking for our work.

- **Oligopolistic Market** is a market where a small number of large sellers dominate. Such market generates opportunity of collusion among these small umber of sellers thereby leading to reduction in competition and higher prices for buyers. Most of real-world markets are oligopolistic [55]. In our work, we consider oligopolistic market to investigate efficiency of our model in realistic settings. We consider oligopolistic market as a representative of real world markets which consist of strategic benefit anticipating players.

- **Price of Anarchy** (PoA) is a concept in economics and game theory that measures degradation in efficiency of a system due to selfish behavior of its agents [19]. In our work, PoA characterizes the efficiency of strategic allocation of privacy compromise amounts by different agents (data holders) at PCE and ONE.

- **VCG Auction** is also known as VickreyClarkeGroves auction. It is is a type of sealed-bid auction of multiple items where bidders submit bids which represents their true valuations for the items, without knowing the bids of the other bidders. The auction system assigns the items in a socially optimal manner: it charges each individual the harm they cause to other bidders. It gives bidders an incentive to bid their true valuations, by ensuring that the optimal strategy for each bidder is to bid their true valuations of the items [57].

- **Social Welfare Function** is a concept from welfare economics which assigns priority to different social states. Such state is assumed to be the description of the society [58]. We discuss three widely used social welfare functions:

    1. **Utilitarian Function** is measured by summing utility number of all members of the society. It promotes the idea of greatest amount of good for the greatest number of people [59].
    2. **Egalitarian Function** maximizes the utility number of the most unfortunate members of the society. It promotes the idea which provides the greatest welfare subject to the constraint that all individual members should enjoy equal benefits from the society [59].
    3. **Rawlsian Function** measures the social welfare of the society on the basis of the welfare of the least well-off individual member of society [59].

Why Choose an Utilitarian Function ? - We choose to work with the utilitarian function over two other popular Bergson-Samuelson social welfare functions used in economic applications: the *egalitarian function*, and the *Rawl's function*, for the following reasons:

- The parameters corresponding to the unique optimal solution of the maximum utilitarian social welfare problem *coincide* with those obtained at the unique equilibrium of a purely distributed market comprising autonomous privacy compromising DH's *without* the presence of a regulator (e.g., ad-network), and are Pareto optimal. This result is due to Arrow-Debreu's first and second fundamental theorems of welfare

economics [55]. In addition, at market equilibrium, there is equitability in the marginal utilities of all the autonomous DHs (in case of DHs, the utility is represented by cost and is thus a negative utility). The parameter coincidence property does not necessarily hold for non-utilitarian social welfare functions.

- The Rawl's social welfare function focuses on maximizing the minimum resource/utility allocation to any stakeholder (e.g., DH in our work). A major drawback of adopting this social welfare function is that it will in general discourage DHs from compromising privacy (even at Pareto optimal system settings), thereby challenging the core philosophy behind a DH market, and will not likely be popular with either the DHs or the regulator (e.g., the ad-network in our work). A max-min utility allocation among DH would favor, for example, a regime that reduces every DH to complete "misery" if it promotes the well-being of the most "miserable" DH by even a very small amount.

- The egalitarian social welfare function focuses on equalizing the utilities of all market stakeholders in the absolute sense. Similar to the case of Rawl's function, it suffers from the major drawback that it will in general discourage DHs from compromising (even at Pareto optimal system settings) privacy. Likewise, it is unlikely to be popular amongst either the regulator or autonomous DHs. For example, if we had to choose between two allocation policies, one under which all DHs would have a cardinal utility of 100, but one DH would have a utility of 99; the second policy under which every DH is "miserable" and will have a cardinal utility of 1 unit. The egalitarian regulator would prefer the latter because under this option, every DH has exactly the same utility level.

B Proofs

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

$$(C_i^{'}(q_i(\bar{b}_i, \bar{p}_i)) - \bar{p})(b_i - \bar{b}_i) \geq 0,\ \forall b_i \geq 0 \tag{20a}$$

$$\sum_i q_i(\bar{b}_i, \bar{p}) = d \tag{20b}$$

Here, (20a) results from the optimality condition of the convex optimization problem of DH net revenue, and (20b) follows directly from Definition 1. Since $\bar{p} \geq 0$, multiplying $\bar{p}$ to (20a), we get

$$\left(C_i^{'}\left(\bar{q}_i\right) - \bar{p}\right)\left(q_i - \bar{q}_i\right) \geq 0,\ \forall q_i \geq 0 \tag{21a}$$

$$\sum_i \bar{q}_i = d \tag{21b}$$

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 (20a) ; 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 (21a), $\bar{p} = C_i^{'}(\bar{q}_i)$ 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 (21a) and (21b). From (21a), 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 $1, 2, ..., \bar{n}$. 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: Theorem 2, we known 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

$$U_i(b_i, b_{-i}) = p(b)q_i(p(b), b_i) - C_i(q_i(p(b), b_i)) = \frac{d^2 b_i}{\left(\sum_j b_j\right)^2} - C_i\left(\frac{db_i}{\sum_j b_j}\right) \tag{22}$$

From (22), we have

$$\begin{aligned}\frac{\partial U_i(b_i, b_{-i})}{\partial b_i} &= \frac{d^2\left(B_{-i} - b_i\right)}{\left(B_{-i} + b_i\right)^3} - \frac{dB_{-i}}{\left(B_{-i} + b_i\right)^2} C_i^{'}\left(\frac{db_i}{B_{-i} + b_i}\right) \\ &= \frac{d^2}{(B_{-i} + b_i)^2}\left[\frac{B_{-i} - b_i}{B_{-i} + b_i} - \frac{B_{-i}}{d} C_i^{'}\left(\frac{db_i}{B_{-i} + b_i}\right)\right]\end{aligned} \tag{23}$$

The first form in the square bracket in (23) 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^*,

$$b^* satisfies \begin{cases} b_i^* = 0, \forall i, \text{if } \frac{B_{-i}^*}{dC_i^{'}(0)} \geq 1 \\ \frac{B_{-i}^* - b_i^*}{B_{-i}^* + b_i^*} - \frac{B_{-i}^*}{d} C_i^{'}\left(\frac{db_i^*}{B_{-i}^* + b_i^*}\right) = 0, \text{otherwise} \end{cases} \tag{24}$$

Given a Nash Equilibrium, b^*: 1) if $b_i^* = 0$, then $b_i^* < B_{-i}^*$ from lemma 1 and, 2) otherwise, b_i^* satisfies (24). Note that the second term on the left hand side of (24) 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: $\min_{0\le q_i<\hat{d}} \sum_i D_i(q_i)$ subject to $\sum_i q_i = d$. Denote optimal value of this problem as $D^*_{\hat{d}}$. For each i, find ε_i such that $D_i(q_i) \ge D^*_{\hat{d}}$ 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: $\min_{0\le q_i\le \frac{d}{2}-\varepsilon_i} \sum_i D_i(q_i)$ 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

$$D_i^{'}(q_i) = \left(1 + \frac{q_i}{d-2q_i}\right) C_i^{'}(q_i) \tag{25}$$

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 $D_i(q_i) = \int_0^{q_i} D_i^{'}(x_i)dx_i \ge C_i^{'}(0)\int_0^{q_i}\left(1+\frac{x_i}{d}-2x_i\right)dx_i = C_i^{'}(0)\int_0^{q_i}(\frac{1}{2}+\frac{d}{2d}-2x_i)dx_i = C_i^{'}(0)\int_0^{q_i}\left(\frac{1}{2q_i}-\frac{d}{4log(d-2x_i)}\right)dx_i$. 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

$$\left(p^* - \left(1+\frac{q_i^*}{d-2q_i^*}\right) C_i^{'}(q_i^*)\right)(q_i - q_i^*) \le 0, \forall q_i \tag{26a}$$

$$\sum_i q_i^* = d \tag{26b}$$

$$p^* > 0 \tag{26c}$$

$$\left(\frac{d}{B_{-i}^* + b_i^*} - \frac{B_{-i}^*}{B_{-i}^* - b_i^*} C_i^{'}\left(\frac{db_i^*}{B_{-i}^* + b_i^*}\right)\right)(b_i - b_i^*) \le 0, \forall b_i \tag{26d}$$

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 (26d) as $\left(p^* - \left(\frac{q_i^*}{d-2q_i^*}\right) C_i^*(q_i^*)\right)(b_i p^* - q_i^*) \le 0$. 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 (26a) - (26c), 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 $D_i^{'}(q_i) \ge C_i^{'}(q_i)$ for any $q_i \in \left[0, \frac{d}{2}\right)$. For any $i \in N$, $(D_i^{'})^{-1}(\bar{p}) \le C_i^{-1}(\bar{p})$. Suppose $p^* < \bar{p}$. Because $C^0_{n^*} \le p^* \le C^0_{n^*+1}$, $C^0_{\bar{n}} \le \bar{p} \le C^0_{\bar{n}+1}$, and $C_1^0 \le C_2^0 \le \ldots\ldots \le C_n^0$, we have $n^* \le \bar{n}$. Therefore, $\sum_i^{n^*}(D_i^{'})^{-1}(p^*) < \sum_i^{n^*}(D_i^{'})^{-1}(\bar{p}) \le \sum_i^{n^*}(C_i^{'})^{-1}(\bar{p}) \le \sum_i^{\bar{n}}(C_i^{'})^{-1}(\bar{p}) = d$, which contradicts that $\sum_i^{n^*}(D_i^{'})^{-1}(p^*) = d$. Thus, $p^* \le \bar{p}$. Therefore, $\bar{n} \le n^*$, implying $\bar{N} \subset N^*$. If $n^* < n$, then $p^* \le D^{'}_{n^*+1}(0) \le D^{'}_{n^*+1}\left(\frac{d}{n}\right) = \frac{n-1}{n-2} C^{'}_{n^*+1}\left(\frac{d}{n}\right) \le$

$\frac{n-1}{n} - 2M$. 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,

$$p^* \leq \frac{n-1}{n-2}M \tag{27}$$

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,

$$\bar{p} \geq C_j'\left(\frac{d}{n}\right) \geq m \tag{28}$$

Combing (27) and (28) 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 $q_i < \frac{d}{2}$, $D_i(q_i) = \left(1 + \frac{q_i}{d-2q_i}\right) C_i(q_i) - \int_0^{q_i} \frac{d}{(d-2x_i)^2} C_i(x_i)dx_i \geq \left(1 + \frac{q_i}{d-2q_i}\right) C_i(q_i) - C_i(q_i) \int_0^{q_i} \frac{d}{(d-2x_i)^2} dx_i \geq \left(1 + \frac{q_i}{d-2q_i}\right) C_i(q_i) - C_i(q_i)\frac{q_i}{d-2q_i} \geq C_i(q_i)$.
Thus, $C^* = \sum_i C_i(q_i^*) \leq \sum_i D_i(q_i^*) \leq \left(1 + \frac{\bar{q}_{max}}{d-2\bar{q}_{max}}\right)\bar{C}$. ■

Assumption 1: For each DH_i, cost function $c_i : (-\infty, \infty) \to (-\infty, \infty)$ is continuous and convex with $c_i(0) = 0$. $C_i()$ is strictly increasing over $[0, \infty)$ over the domain $(-\infty, 0)$, $c_i()$ is either (i) non-negative or (ii) strictly increasing on $[-L_i, 0)$ and equals $c_i(-L_i)$ in $(-\infty, -L_i)$, $L_i \geq 0$ is a constant.

Assumption 2: For every i, $\Sigma_{j\neq i}^n D_j > d$.

Proof of Theorem 6: Given $p > 0$, each DHs' payoff function is concave. Thus, an action vector is a competitive equilibrium, if and only if each of its components $b_i \in [0, p(D_i + L_i)]$, and satisfies the following condition:

$$\frac{\delta^- c_i(S_i(b_i, p))}{\delta q_i} \leq p,\ 0 \leq b_i < p(D_i + L_i) \tag{29a}$$

$$\frac{\delta^+ c_i(S_i(b_i, p))}{\delta q_i} \geq p,\ 0 < b_i < p(D_i + L_i) \tag{29b}$$

Note that, at a competitive equilibrium b, DH_i would never submit a bid that is larger than $p(D_i + Li)$, in that case, DH_i obtains a negative payoff because $S_i(b_i, p) < -L_i$. Since objective function in OPT_1 is convex, and the optimization problem is over a convex, compact feasible set, there exists an optimal solution. The following condition is necessary and sufficient for a feasible solution to OPT_1, $q^\star$ to be optimal:

$$\frac{\partial^- c_i(q_i^\star)}{\partial q_i} \leq \mu,\ -L_i < q_i^\star \leq D_i \tag{30a}$$

$$\frac{\partial^- c_i(q_i^\star)}{\partial q_i} \geq \mu,\ -L_i \leq q_i^\star < D_i \tag{30b}$$

Here, μ is the Lagrange multiplier for the constraint $\Sigma_i q_i = d$. Since c_i is strictly increasing over $[-L_i, D_i]$ and at least one component of $q^\star$ is positive , it follows that

$\mu > 0$. It is not hard to see that the action vector $\{b_i = \mu(D_i - q_i^\star)\}$ with $p = \mu$ satisfies 29, and is therefore a competitive equilibrium. On the other hand, for any competitive equilibrium, it is straightforward to show that the supply vector $\{D_i - \frac{b_i}{p}\}_{i=1}^n$, together with a μ value set equal to the point $p > 0$ satisfies the condition 30, i.e., the resulting allocation is socially optimal. ■

Proof of Lemma 4: In order to clear market, we have

$$\Sigma_{i=1}^n S_i(b_i, p) = \Sigma_{i=1}^n (D_i - \frac{b_i}{p}) = d \tag{31}$$

The market clearing benefit is given by

$$p = \frac{\Sigma_{i=1}^n b_i}{-d + \Sigma_{i=1}^n D_i} \geq 0 \tag{32}$$

Thus, the supply provided by DH_i a function of b is given as:

$$q_i = D_i - \frac{b_i(-d + \Sigma_j D_j)}{\Sigma_j b_j} \tag{33}$$

Note that q_i decreases and converges to $d + D_i - \Sigma_j D_j > 0$, as b_i increases to ∞. On the other hand, the market clearing benefit p blows up as b_i grows larger. This DH_i's payoff is unbounded as her bid b_i increases to ∞. A Nash equilibrium does not exist. ■

Proof of Lemma 5: Suppose that b is a Nash equilibrium such that $b_j = 0$ for every $j \neq i$. DH_i's payoff is given by

$$\Pi_i(b_i, 0) = \begin{cases} \frac{D_i b_i}{-d + \Sigma_{j=1}^n D_j} - b_i \\ -c_i(D_i + d - \Sigma_{j=1}^n D_j), \text{if } b_i > 0 \\ -c_i(\frac{dD_i}{\Sigma_j D_j}), \text{if } b_i = 0 \end{cases} \tag{34}$$

According to Assumption 2, we have

$$\Sigma_j D_j - d > D_i \tag{35}$$

It follows that DH_i would like to submit an arbitrarily low bid to minimize the market clearing benefits. The vector b cannot be a Nash equilibrium. ■

Proof of Theorem 7: We first derive necessary and sufficient conditions for an action vector b to be a Nash equilibrium. We then show that there exists a unique optimal solution in OPT_2. We then derive necessary and sufficient optimally condition for the optimization problem in OPT_2. The correspondence between the Nash equilibrium condition establishes the existence of a Nash equilibrium and the uniqueness of the resulting allocation.

Step 1. (Necessary and Sufficient Nash equilibrium condition): We argue in this step that a vector b is a Nash equilibrium, if and only if it has at least two positive components, each of its components $b_i \leq c_i$, (where c_i is a constant to be defined below), and it satisfies the following condition:

$$\frac{\partial^- c_i(S_i(b_i, p(b)))}{\partial q_i} \left(1 + \frac{S_i(b_i, p(b))}{-d + \Sigma_{j \neq i} D_j}\right) \leq p(b),\ 0 \leq b_i < c_i \tag{36a}$$

$$\frac{\partial^+ c_i(S_i(b_i, p(b)))}{\partial q_i} \left(1 + \frac{S_i(b_i, p(b))}{-d + \Sigma_{j \neq i} D_j}\right) \geq p(b),\ 0 < b_i \leq c_i \tag{36b}$$

where

$$c_i = \begin{cases} \frac{(D_i+L_i)\Sigma_{j\neq i}b_j}{-d-L_i+\Sigma_{j\neq i}D_j}, \text{if } -d-L_i+\Sigma_{j\neq i}D_j > 0 \\ \infty, \text{if } -d-L_i+\Sigma_{j\neq i}D_j < 0 \end{cases} \tag{37}$$

We note that, if $-d - L_i + \Sigma_{j\neq i}D_j > 0$, then at Nash equilibrium b, DH_i would never submit a bid that is larger than

$$e_i = \frac{(D_i + L_i)\Sigma_{j\neq i}b_j}{-d - L_i + \Sigma_{j\neq i}D_j} \tag{38}$$

This is because bidding c_i will always yield DH_i a higher payoff than submitting a bid larger than c_i. On the other hand, if $-d - L_i + \Sigma_{j\neq i}D_j \leq 0$, we have $S_i(b_i, p(b)) \geq -L_i$ under all possible non-negative vector b.

Let b be a Nash equilibrium based on lemma 2, we have at least two components of b are positive, and the market clearing benefit $p(b)$ is positive. This DH_i's payoff is given by

$$\Pi_i(b_i, b_{-i}) = \frac{D_i(b_i + \Sigma_{j\neq i}b_j)}{-d + \Sigma_{j=1}^n D_i} - b_i - c_i\left(D_i - \frac{b_i(-d + \Sigma_{j=1}^n D_j)}{b_i + \Sigma_{j\neq i}b_j}\right) \tag{39}$$

which can be shown to be continuous and concave in b_i, over the domain $(0, \infty)$. Since for every i, $\Pi(b_i, b_{-i})$ is concave in b_i, the following condition is necessary and sufficient for a vector b to be a Nash equilibrium.

$$\frac{\partial^+\Pi_i(b_i, b_{-i})}{\partial b_i} \leq 0,\ 0 \leq b_i < c_i \tag{40a}$$

$$\frac{\partial^-\Pi_i(b_i, b_{-i})}{\partial b_i} > 0,\ 0 < b_i \leq c_i \tag{40b}$$

Substituting 39 in 40, we get

$$\frac{\partial^- c_i(S_i(b_i, p(b)))}{\partial q_i}\left(1 - \frac{b_i}{\Sigma_j b_j}\right) \leq \frac{-d + \Sigma_{j\neq i}D_j}{-d + \Sigma_j D_j}p(b),\ 0 \leq b_i < c \tag{41a}$$

$$\frac{\partial^+ c_i(S_i(b_i, p(b)))}{\partial q_i}\left(1 - \frac{b_i}{\Sigma_j b_j}\right) \leq \frac{-d + \Sigma_{j\neq i}D_j}{-d + \Sigma_j D_j}p(b),\ 0 < b_i \leq c \tag{41b}$$

Through the following relation:

$$1 - \frac{b_i}{\Sigma_j b_j} = 1 - \frac{(D_i - S_i(b_i, p(b)))p(b)}{(\Sigma_j D_j - \Sigma_j S_j(b_j, p(b)))p(b)} = \frac{-d + \Sigma_{j\neq i}D_j + S_i(b_i, p(b))}{\Sigma_j D_j - d} \tag{42}$$

It is easy to see that 41 is equivalent to 36.
Step 2: (Existence and Uniqueness of an optimal solution to OPT_1)
We show that there exists a unique optimal solution to OPT_1. We first argue that $\hat{c}_i(q_i)$ is continuous, strictly convex, and strictly increasing over $q \geq -L_i$, where

$$\hat{c}_i(q_i) = \begin{cases} \left(1 + \frac{q_i}{-d+\Sigma_{j\neq i}D_j}\right)c_i(q_i) - \frac{1}{-d+\Sigma_{j\neq i}D_j}\int_0^{q_i} c_i(x)dx, \text{if } q_i \geq 0 \\ \left(1 + \frac{q_i}{-d+\Sigma_{j\neq i}D_j}\right)c_i(q_i) - \frac{1}{-d+\Sigma_{j\neq i}D_j}\int_{q_i}^{0} c_i(x)dx, \text{if } q_i < 0 \end{cases} \tag{43}$$

From 43, we have

$$\frac{\partial^- \hat{c}_i(q_i)}{\partial q_i} = \left(1 + \frac{q_i}{-d + \Sigma_{j \neq i} D_j}\right) \frac{\partial^- c_i(q_i)}{\partial q_i} \tag{44a}$$

$$\frac{\partial^+ \hat{c}_i(q_i)}{\partial q_i} = \left(1 + \frac{q_i}{-d + \Sigma_{j \neq i} D_j}\right) \frac{\partial^+ c_i(q_i)}{\partial q_i} \tag{44b}$$

$$\tag{44c}$$

Suppose c_i is strictly increasing and convex, for any $-L_i \leq q_i < q_i^{'}$, we have

$$0 \leq \frac{\partial^+ \hat{c}_i(q_i)}{\partial q_i} < \frac{\partial^- \hat{c}_i(q_i^{'})}{\partial q_i^{'}} \leq \frac{\partial^+ \hat{c}_i(q_i^{'})}{\partial q_i^{'}} \tag{45}$$

This implies that $\hat{c}_i$ is strictly increasing and strictly convex over $[-L_i, \infty]$. Since, for every i, $\hat{c}_i$ is continuously and strictly convex, OPT_1, over a convex, compact feasible region must have a unique optimal solution.
Step 3: (Necessary and Sufficient Optimally condition for OPT_1
Let $q = (q_1, q_2, ..., q_n)$ be the unique optimal solution to OPT_i. There exists a Lagrange multiplier μ such that

$$\left(1 + \frac{q_i}{-d + \Sigma_{j \neq i} D_j}\right) \frac{\partial^- c_i(q_i)}{\partial q_i} \leq \mu, \ -L_i < q_i \leq D_i \tag{46a}$$

$$\left(1 + \frac{q_i}{-d + \Sigma_{j \neq i} D_j}\right) \frac{\partial^+ c_i(q_i)}{\partial q_i} \leq \mu, \ -L_i \leq q_i < D_i \tag{46b}$$

Since at least one q_i is positive and c_i is strictly increasing, we have $\mu > 0$. We now consider that action vector $\{b_i = (D_i - q_i)u\}_{i=1}^n$. Note that at least two components of b are positive because $\Sigma_{j \neq i} D_j > d$ for every i. Since $q_i = D_i$, if and only if $b_i = 0$, and $q_i = -L_i$ if and only if $b_i = c_i$ it is not hard to see from 46 that the action vector $\{(D_i - q_i)u\}_{i=1}^n$ satisfies condition in 36, and is therefore a Nash equilibrium.

Finally, we argue that all Nash equilibria result in the same privacy compromise that is an optimal solution to OPT_1. A Nash equilibrium b satisfies condition in 36. It follows that the vector $\{S_i(b_i, p(b))\}_{i=1}^n$ satisfies condition in 46, with $p(b) > 0$ being the Lagrange multiplier. Since $\hat{c}_i$ is strictly convex for every i, condition in 36 suffice that $\{S_i(b_i, p(b))\}_{i=1}^n$ is an optimal solution to OPT_1.

Step 4: (Uniqueness of Nash equilibrium Under an Additional Assumption)
We are left to show that the uniqueness of Nash equilibrium under an additional assumption that all DHs have continuously differentiable cost function. In this setting, the necessary and sufficient Nash equilibrium condition 36 can written as

$$c_i^{'}(S_i(b_i, p(b)) \left(1 + \frac{S_i(b_i, p(b))}{-d + \Sigma_{j \neq i} D_j}\right) \leq p(b), \text{if } 0 \leq b_i < c_i \tag{47a}$$

$$c_i^{'}(S_i(b_i, p(b)) \left(1 + \frac{S_i(b_i, p(b))}{-d + \Sigma_{j \neq i} D_j}\right) \geq p(b), \text{if } 0 < b_i \leq c_i \tag{47b}$$

Suppose that there are two distinct Nash equilibria b and $b^{'}$, we now prove the uniqueness of Nash equilibrium by considering:
1) If there exists a DH_i such that $b_i \in (0, c_i)$ at the Nash equilibrium b, then according to 47, we have

$$c_i^{'}(S_i(b_i, p(b))) \left(1 + \frac{S_i(b_i, p(b))}{-d + \Sigma_{j \neq i} D_j}\right) = p(b) \tag{48}$$

We also note that DH_i's compromise $S_i(b_1, p(b))$ must lie in the interval $(-L_i, D_i)$. We have shown that DH_i must provide the same amount of privacy compromise at the two Nash equilibrium b and $b^{'}$, and as a result, $S_i(b_i^{'}, p(b)) \in (-L_i, D_i)$. We thus have

$$c_i^{'}(S_i(b_i^{'}, p(b^{'}))) \left(1 + \frac{S_i(b_i^{'}, p(b^{'}))}{-d + \Sigma_{j \neq i} D_j}\right) = p(b^{'}) \tag{49}$$

The above two equalities lead to $p(b) = p(b^{'})$, which in turn implies $b = b^{'}$.

2) Suppose now that there does not exist a DH_i such that $b_i \in (0, c_i)$. At the Nash equilibrium b, each DH_i bids either c_i or 0, and provides either $-L_i$ or D_i privacy compromise. In this case the n DHs can be divided into two groups A and B, such that every DH in group A bids c_i and every DH in group B bids 0. From 47 we have

$$c_j^{'}(D_j) \left(1 + \frac{D_j}{-d + \Sigma_{k \neq j} D_k}\right) \leq p(b) \leq c_i^{'}(-L_i) \left(1 + \frac{-c_i}{-d + \Sigma_{k \neq i} D_k}\right) \tag{50}$$

for every $i \in A$ and every $j \in B$. It is straightforward to check (prove 47) that at any Nash equilibrium, the benefit must lie in the following range:

$$\left(\max_{j \in B} c_j^{'} \frac{-d + \Sigma_k D_k}{-d + \Sigma_{k \neq j} D_k}, \min_{i \in A} c_i^{'} \frac{-d - L_i + \Sigma_{k \neq i} D_k}{-d + \Sigma_{k \neq i} D_k}\right) \tag{51}$$

Otherwise the total supply would not be d. it follows that b is the unique Nash equilibrium.■

Proof of Corollary 4: We note that DH_i achieves a zero payoff if his compromise adjustment is zero. This implies that a DH must obtain at least a zero payoff at a Nash equilibrium, because given any bids submitted by other DHs, a DH can always choose a bid that yields zero supply. Let b be a Nash equilibrium. Suppose that DH_i's supply is positive at the equilibrium, i.e., $S_i(b_i, p(b)) > 0$. It follows from 36 that

$$\frac{\partial^{-} c_i(S_i(b_i, p(b)))}{\partial q_i} < p(b) \tag{52}$$

Since c_i is a convex function, we have

$$0 < c_i(S_i(b_i, p(b))) \leq \frac{\partial^{-} c_i(S_i(b_i, p(b)))}{\partial q_i} S_i(b_1, p(b)) < p(b) S_i(b_1, p(b)) \tag{53}$$

which implies that DH_i achieves a positive payoff at the equilibrium. On the other hand if $S_i(b_i, p(b)) < 0$ it follows from 36 that

$$\frac{\partial^{+} c_i(S_i(b_i, p(b)))}{\partial q_i} > p(b) \tag{54}$$

We similarly have

$$c_i(S_i(b_i, p(b))) \leq \frac{\partial^{+} c_i(S_i(b_1, p(b)))}{\partial q_i} S_i(b_i, p(b)) < p(b) S_i(b_1, p(b)) < 0 \tag{55}$$

■

Proof of Corollary 5: Since $\{q_i\}_{j=1}^n$ is he supply vector corresponding to a Nash equilibrium with the capacity limit $\{D_i\}_{j=1}^n$ and $q_i < D_i$, it follows from 47 that

$$\left(1 + \frac{q_i}{-d + \Sigma_{j \neq i} D_j}\right) \frac{\partial c_i(q_i)}{\partial q_i} \geq \mu \tag{56}$$

We will show that proposition by contradiction. Suppose that $\bar{q}_i > q_i$. Since $\{\bar{q}_j\}_{j=1}^n$ is the supply vector corresponding to a Nash equilibrium with the capacity limit $\{\bar{D}_j\}_{j=1}^n$ and $-L_i \leq q_i < \bar{q}_i$, it follows from 47 that

$$\left(1 + \frac{\bar{q}_i}{-d + \Sigma_{j \neq i} \bar{D}_j}\right) \frac{\partial c_i(\bar{q}_i)}{\partial \bar{q}_i} \leq \bar{\mu} \tag{57}$$

Since $D_j = \bar{D}_j$ for every $j \neq i$, $\bar{q}_i > q_i$, we must have $\bar{\mu} > \mu$. Since $\Sigma_{j=1}^n q_j = \Sigma_{j=1}^n \bar{q}_j$, there must exist some DH k such that $\bar{q}_k < q_k$. Since $-L_k \leq \bar{d}_k < d_k \leq D_k$, it follows from 47 that

$$\left(1 + \frac{q_k}{-d + \Sigma_{j \neq k} D_j}\right) \frac{\partial c_k(q_k)}{\partial q_k} \leq \mu \tag{58a}$$

$$\left(1 + \frac{\bar{q}_k}{-d + \Sigma_{j \neq k} \bar{D}_j}\right) \frac{\partial c_k(\bar{q}_k)}{\partial \bar{q}_k} \leq \bar{\mu} \tag{58b}$$

Since $\Sigma_{j \neq k} D_j < \Sigma_{j \neq k} \bar{D}_j$ and $q_k > \bar{q}_k$, the above two inequalities imply that $\bar{\mu} < \mu$. Thus, we conclude that $\bar{q}_i \leq q_i$. ■

Proof of Theorem 8: We first show that the virtual cost function $\hat{c}_j$ is larger than original cost function c_j. Since c_j is non-decreasing for $j = 1, ...n$ and $q \in [0, min\{D, D_j\}]$, we have

$$\hat{c}_j(q) \geq \left(1 + \frac{q}{-d + \Sigma_{k \neq j} D_k}\right) c_j(q) - \frac{1}{-d + \Sigma_{k \neq j} D_k} \int_0^q c_j(x) dx = c_j(q) \tag{59}$$

Similarly, for $j = 1, ...n$ and $q \in [-L_j, 0)$, we have

$$\hat{c}_j(q) \geq \left(1 + \frac{q}{-d + \Sigma_{k \neq j} D_k}\right) c_j(q) + \frac{1}{-d + \Sigma_{k \neq j} D_k} \int_q^0 c_j(x) dx = c_j(q) \tag{60}$$

On the other hand, $c_j(q) \geq 0$ for $q \geq 0$, we have

$$\hat{c}_j(q) \leq \left(1 + \frac{q}{-d + \Sigma_{k \neq j} D_k}\right) c_j(q), \ q \geq 0 \tag{61}$$

It follows that, for $j = 1, ...n$ and $q \in [0, min\{d, D_j\}]$ We have

$$\hat{c}_j(q) \leq \left(1 + \frac{min\{D_j, D\}}{-d + \Sigma_{k \neq j} D_k}\right) c_j(q) \tag{62}$$

Let $q^\star$ be the non-negative socially optimal allocation and q be an allocation resulting from a Nash equilibrium respectively. We have

$$\begin{aligned} \Sigma_{j=1}^n c_j(q_j) \leq \Sigma_{j=1}^n \hat{c}_j(q_j) \leq \Sigma_{j=1}^n c_j^\star(q_j^\star) &\leq \Sigma_{j=1}^n \left(1 + \frac{min\{D_j, d\}}{-d + \Sigma_{k \neq j} D_k}\right) c_j(q_j^\star) \\ &\leq \left(1 + \frac{min\{D_i, d\}}{-d + \Sigma_{k \neq i} D_k}\right) \Sigma_{j=1}^n c_j(q_j^\star) \end{aligned} \tag{63}$$

Here, the first inequality is true because $\hat{c}_j(q_j) \leq c_j(q_j)$ for every j and every q_j, the second inequality follows from the fact that q minimizes the sum of virtual cost function $\hat{c}_j$, the third inequality follows from 62, and the last inequality is true because $D_i = max_j\{D_j\}$.

It remains to be shown that the bound is higher for these $d \leq D_i$. Fixing $d > 0$ and $n \geq 2$, we consider a model where $D_1 \geq D_2 = ... = D_n$ and $L_j = 0$ for very j. Let r be positive constant such that $\frac{d}{n} < r \leq D_i$ and let $\delta \in (0,1)$ DH_i's cost function is

$$\hat{c_1}(q_1) = \begin{cases} \delta q_1, \text{if } 0 \leq q_1 \leq r \\ q_1 - r + \delta r, \text{if } r < q_1 \leq D_1 \end{cases} \tag{64}$$

and for $j = 2, ..., n$, $c_j(q_j) = 2q_j, q_j \geq 0$, where

$$\alpha = \frac{1 + \frac{r}{-d + \Sigma_{k=2}^{n} D_k}}{1 + \frac{d-r}{(-d+\Sigma_{k\neq 2}^{n} D_k)(n-1)}} \tag{65}$$

Since $\frac{d}{n} < r$ and $D_1 = max_k\{D_k\}$, it follows that $\alpha > 1$. Thus, a socially optimally allocation is given by 1) $q_1^\star = d_j$, 2) $q_j^\star = 0$, $j \geq 2$
We now argue that the supply vector $q = (q_1, ...q_n) = \left(r, \frac{d-r}{n-1}, ..., \frac{d-r}{n-1}\right)$ is an optimal solution to OPT_1. To see this, let $\mu = 1 + \frac{r}{-d+\Sigma_{k=2}^{n} D_k}$, and we have

$$\left(1 + \frac{q_1}{-d + \Sigma_{k\neq 1} D_k}\right) \frac{\partial^- c_1(q_1)}{\partial q_1} = \delta\mu \leq \mu_i \tag{66a}$$

$$\left(1 + \frac{q_1}{-d + \Sigma_{k\neq 1} D_k}\right) \frac{\partial^+ c_1(q_1)}{\partial q_1} = \mu_i \tag{66b}$$

$$\left(1 + \frac{q_j}{-d + \Sigma_{k\neq 1} D_k}\right) \frac{\partial c_j(q_j)}{\partial q_j} = \mu_i, j = 2, ..., n \tag{66c}$$

Since the preceeding condition is equivalent to 47, it follows that q is an optimal solution to OPT_1, and is therefore the allocation result form a Nash equilibrium. At the Nash equilibrium, the aggregate utility loss is $\Sigma_j c_j(q_j) = \delta r + \alpha(d-r)$, while at social optimal we have

$$\Sigma_j c_j(q_j^\star) = d - r + \delta r \tag{67}$$

We obtain

$$\frac{\Sigma_j c_j q_j}{\Sigma_j c_j(q_j^\star)} = \frac{\delta r + \alpha(d-r)}{d - r + \delta r} \tag{68}$$

Let $r \to d$ and $\frac{\delta r}{d-r} \to 0$ e.g., $\delta = (d-r)^2$. The preceding ration converges to α, whose limit is given by

$$\lim_{r\to d} \frac{1 + \frac{r}{-d+\Sigma_{k=2}^{n} D_k}}{\frac{d-r}{\left(-d+\Sigma_{k\neq 2}^{n} D_k\right)(n-1)}} = 1 + \frac{d}{-d + \Sigma_{k=2}^{n} D_k} \tag{69}$$

■

Proof of Corollary 6: Since $S_i(b_i, p(b)) < D_i$, we have $b_i > 0$. We also know that

$$\frac{\partial^+(S_i(b_i, p(b)))}{\partial q_i} \geq \frac{\partial^+(S_i(b_i, p(b)))}{\partial q_i}\left(1 - \frac{b_j}{\Sigma_j b_j}\right) \geq \frac{-d + \Sigma_{j\neq i} D_j}{-d + \Sigma_j D_j} p(b) \tag{70}$$

Thus, we have yielded the bound. ■