A Negotiation Game: Establishing Stable Privacy Policies for Aggregate Reasoning

Rosa Karimi Adl, Ken Barker, and Jörg Denzinger

Department of Computer Science, University of Calgary, Canada
{rkarimia,kbarker}@ucalgary.ca{denzinge}@cpsc.ucalgary.ca

Abstract. The process of personal information collection and exchange is associated with ever-growing privacy concerns. To resolve the issue, data provider's consent on the usage of private information is sought through privacy policy specifications. The parameters of such privacy policies influence the quantity and quality of gathered information. Choosing the right privacy policy parameters can potentially increase the revenues to a data collector and the firms (third-parties) interested in accessing the database for data analysis purposes. In this work we use an extensive form game model to examine the decisions made by a data collector and a third-party to maximize their benefits from collecting and accessing data. We have found the game's subgame perfect equilibria for various problem settings and provide the details of game analysis for a simplified scenario and two case studies. The equilibrium solutions demonstrate steady states of the game where collecting personal information at a specific privacy level is advantageous to the data collector and the third-party. Consequently the results define a realistic boundary on collecting personal information.

Keywords: Data Privacy, Data Repository Management, Privacy Policy Setting, Price/Privacy Trade-off, Game Theory

1 Introduction

Todays ever increasing privacy concerns stem from advances in mass data storage technologies, Web-mediated data collection/access methods, and data mining procedures to discover hidden patterns. The ideal of having perfect privacy protection has been shown to be very ambitious [11]. Even if a sanitization method [24,15,12] (which aims at creating a publishable private data table) is used, the amount of privacy protection is limited by the value of a privacy parameter (such as k in k-anonymity). Regardless of the privacy protection mechanism, *some* risk to an individual's privacy is still present. Therefore, data providers' consent in terms of an agreement to a privacy policy must be sought before collecting personal information.

A privacy policy can be specified using languages such as P3P [9], XACML [18], or EPAL [5]. Specification of a privacy policy involves setting some parameters to define scope and limitations of the promised privacy protection level. As the privacy protection level increases, more data providers will be willing to provide their personal information but the collected data will have lower quality for data analysis purposes. A setting for privacy parameters that balances level of data quality/quantity is desired by different parties involved in an information collection and usage procedure. A comprehensive and fair solution to address this challenge must consider the different and often opposing needs of these parties.

We recognize three parties involved in a data collection procedure: The first entity is a *third-party (data user)* who is interested in accessing a private data table with a considerable number of high quality private data records. The second entity is a *data collector* who has earned trust of individuals to protect their privacy. The data collector is asked by the third-party to collect a private table. Before collecting the required information, the data collector publishes a privacy policy. Finally, *data providers* form the third entity who decide whether or not to agree to a privacy policy and participate in the data collection procedure.

We use a sequential game model to illustrate and analyze the interplay of decisions made by three parties (a third-party, a data collector, and data providers) involved in a data collection procedure. The game's outcomes indicate *steady* (stable) privacy policies which are unlikely to be changed since no single player can achieve a higher utility by unilaterally changing his strategy. These steady privacy policies are determined based on how valuable the collected data is to the third-party and how private it is to the data providers. We solve the game with regard to a specific aggregate query application. Our results show the stable combinations of revelation level (how specific data is revealed), retention period, price per data item, and the incentive required to attract data providers.

The game's equilibria are described as choices between fifteen options. However, for each instance of the problem only some of the options are relevant. Each option is a two variable payoff function that the third-party must optimize subject to some constraints on the variables. The data collector's maximum payoff and the optimum amount of incentive are also associated with every option. Therefore, the outcome of this paper is similar to a lookup table which describes realistic privacy settings and the value of information in various situations. The results are also provided for a simplified scenario through a more in-depth analysis. We further demonstrate how to use the results through two case studies.

The novelty of our work lies in considering the reaction of data providers to a privacy policy and framing the challenge of privacy policy settings in a sequential game model to analyze the strategies of data providers, the data collector, and the third-party simultaneously.

The results of this paper show the effects of data providers' privacy behavior on the amount of profit a data collector and a third-party receive under various privacy policies. This outcome implies that realizing the "limited information collection" principle of Hippocratic databases [3] is case based and the boundary on privacy promises is determined based on several parameters. Moreover, as demonstrated in Section 9 the results can be used as a metric to distinguish between sensitive and non-sensitive attributes.

Before explaining the game-theoretic model, we first provide a brief overview of related work in Section 2 and then Section 3 defines the privacy policy structure considered in this paper. Section 4 explains the assumptions made to analyze the situation. Based on these assumptions, the game model, the players, and their payoff functions are described in Section 5. By applying the backward induction method, best responses of the data collector are explored in Section 6. Section 7 considers the third-party's best strategies and the subgame perfect equilibrium of the game. The Equilibrium strategies are analyzed in more detail for a simplified scenario in Section 8. To illustrate the application of the results, we use two case studies in Section 9 and elaborate the procedure for finding the subgame perfect equilibria of the game. Finally, in Section 10 conclusions are drawn and possible extensions to this work are discussed.

2 Related Work

Literature on privacy protected data repositories often follows two major trends. The first group of researchers [24,15,12] mainly focuses on proposing methods to sanitize a database and remove the possibility of tracing a piece of information to any individual. From this point of view a database can be safely published once it is *sanitized*. The second group [5,3,10,6] believe that once the information collection and usage procedure conforms to an agreement between data providers and the data collector (in terms of a privacy policy) data practices are privacy preserving.

With regard to the first point of view, Dwork [11] proves that perfect privacy is unachievable. As a result, all works in this area introduce a parameter such as k in k-anonymity [24], l in l-diversity [15], or t in t-closeness [12] that needs to be initialized by the data collector and guarantees privacy only to the limit specified by these parameters. Only a few researchers [22,13,14,16] provide directions on how to choose the value of privacy parameter. Even these works overlook the effect of data providers' privacy preferences on a balanced value for the privacy parameter.

The second viewpoint has led to extensive research on privacy policy specification standards and elements. The Platform for Privacy Preferences (P3P) [9], eXtensible Access Control Markup Language (XACML) [18] and the Enterprise Privacy Authorization Language (EPAL) [5] are three of the most well-known protocols to specify general purpose and *adhoc* privacy policies. These protocols together with other work on purpose [23,25] only aim at providing structured vocabularies to describe any arbitrary privacy policy. However, the existing literature only hints at how to find the appropriate level of privacy promises. One very significant work in this area is the collection of ten principles in Hippocratic Databases [3]. Two relevant principles in this regard are "limited collection" and "limited retention" that are not formally defined.

To address the challenges of privacy settings in both trends, we use a sequential game model to consider the conflicting needs of all parties involved in a data collection procedure. In an earlier work [2], we have successfully applied this model to address the problem of privacy parameter settings in sanitization systems (and specially

k-anonymity). This work completes our previous work in showing the applicability of our game model to privacy policy specification approach. Both works share the same definition for players and game structure. But since the context of this paper is not data sanitization, this paper has its own descriptions for actions available to players and conditions on the moves. Consequently the whole game analysis procedure is new and different from our previous work.

Game theory has been applied to privacy related issues mainly to find the impact of privacy on price discrimination [7], comparison between self-regulation procedures versus government interference to preserve privacy [4], and modelling privacy discrimination (dynamic privacy) for different groups of data providers [20]. The model explained in the latter is very similar to ours but the game is only used to provide a visual illustration of the challenge and no further analysis is provided. Calzolari and Pavan [8] also use game theory to optimize the flow of private information between two firms. The model in this work is the closest to ours in terms of the entities they recognized as players. However, the analysis is substantially different from what we do since a privacy policy is defined as the probability of revealing detailed customers' information to another party.

Other economic concepts such as economic price theory have also been used to find the optimum trade-off levels between privacy and utility [26]. Zielinski and Olivier [26] use entropy-based metrics to measure the importance of an attribute value to a data user and a privacy intruder. A weighted sum of these two metrics define a utility function that must be optimized under some restrictions. While the ultimate goal of this work is very similar to ours, the problem formulation and the tools used by Zielinski and Olivier are very different from this paper. Our paper does not measure privacy according to an intruder's opinion. Instead we consider the important role of data providers and the effects of privacy protection levels on the quantity of the collected records. We believe that an assumption about having knowledge on aggregate behaviour of data providers is more realistic than having knowledge about intruders' needs and preferences.

3 Privacy Policy

To describe any arbitrary privacy policy, we adopt a framework close to what P3P [10] and at least one privacy taxonomy [6] suggest. The main idea behind these two protocols is that for each piece of collected information (data field), purpose, visibility (recipient), retention period, and granularity levels must be specified. The granularity level is adopted from the privacy taxonomy [6]. It specifies how specific and accurate the value of a data field would appear in a query result. Since granularity levels defined in the privacy taxonomy are not exactly in the context of a privacy policy specification we provide our own granularity levels:

(0)-None: No information on the data field is provided.
(1)-Unlinkable, Partial: The value of a data field cannot be linked to values of other data fields provided by the same individual (Unlinkable) and the value of the data field is generalized or perturbed with some noise.
(2)-Unlinkable, Exact: The data field is not linkable but the value of the field is revealed in the exact form.
(3)-Linkable, non-identifiable, Partial: the data field is linkable to all other linkable data fields. A sanitization method is used to anonymize data. The exact value of the data is not revealed.
(4)-Linkable, non-identifiable, Exact: It is the same as as level (3), except that data value is exact.
(5)-Linkable, identifiable, Partial: It is the same as level (3) without any sanitization.
(6)-Linkable, identifiable, Exact: It is the same as level (4) without any sanitization.

Let DF, Pr, V, R, and $G = \{0, 1, 2, 3, 4, 5, 6\}$ denote the sets of all possible data fields, purposes, visibilities, retentions, and granularity levels. A privacy statement ps can be defined as follows:

$$ps \in DF \times Pr \times V \times R \times G \tag{1}$$

Consequently a privacy policy PP can be defined as a set of privacy statements:

$$PP \subset DF \times Pr \times V \times R \times G \tag{2}$$

The set of privacy statements is usually chosen by the data collector according to data requirements of third-parties interested in the database. Once the data collector publishes a privacy policy PP, the data providers have the choice of opting in or out for the statements. To provide a semantically consistent functionality of opt-in/opt-out options, we consider statement groups in the sense that if a group of privacy statements share the same purpose and visibility, either all or none of the statements in the group must be accepted. Some privacy policy languages such as P3P [9] already support this idea via the "consent attribute" in the statement group construct.

4 Problem Definition and Basic Assumptions

In the process of collecting personal information, three groups of entities are involved: third-parties, a data collector, and data providers. Third-parties are those entities who want to use a data table containing private information for some data analysis purposes and are willing to pay a data collector to collect such information. Upon receiving such a request from a third-party, a data collector announces a privacy policy and possibly some incentive to attract data providers. Finally, data providers decide to participate in the data collection procedure if they find the combination of the privacy policy and incentive as "worthwhile". Since data providers could potentially be anyone who wishes to use a service on line, we use the terms "data providers" and "public" interchangeably throughout the balance of the paper.

Let k denote the total number of data fields a data collector might want to collect from a target population with n individuals as potential data providers. To define a payoff function for every potential data provider, we must know the underlying privacy preferences of every single individual who might participate in the data collection procedure. Since the details of such a diverse set of privacy preferences are usually not known (even to the data providers themselves), in this paper we assume that for each combination of purpose and visibility, data providers' willingness to share their private information is explained by a probability model of the following form:

$$prob(opt - in) = \beta_0 + \beta_1 f(g_1) + ... + \beta_k f(g_k) + \theta h(r) + \gamma I \quad (3)$$

where g_i denotes the granularity level of the data field df_i, r denotes the retention period in terms of the number of years, and I is a real number representing the amount of incentive (in monetary value) the data collector offers the data providers in exchange for the requested information.

$f(.)$ and $h(.)$ are decreasing functions of granularity and retention. Consequently, all β 's, θ, and γ can be safely assumed to have a value greater than or equal to zero. As the granularity levels (or the retention period) increase, the values of function $g(.)$ (or $h(.)$) decrease and the data providers are less willing to share their information. In this paper we make another simplifying assumption and assume that these two functions are defined as:

$$f(g) = \frac{1}{g+1} \; and \; h(r) = \frac{1}{r+1} \tag{4}$$

If functions $f(.)$ and $h(.)$ have definitions different from what we assumed in Eq(4) the procedure explained in this paper will remain the same but the final results may vary.

Various studies [1][17][21] have already provided models and statistics to quantify the influences of factors such as the amount of knowledge about privacy risks, trust, age, income level, *etc.* on the data provider's privacy decisions. A probability model similar to Eq(3) is a natural extension of the existing literature.

The probability model can be viewed as data providers' mixed strategy presumably found based on some observation (and not game analysis). Although such an assumption might seem as an oversimplification of the problem, it increases the applicability and flexibility of our model, encapsulates the social intricacies of human mind, and allows our approach to adopt to any potential evolution in in public's privacy awareness. Nevertheless, the impact of data providers' decisions are still reflected in our model and it is very essential to our game analysis.

A data collector decides on a privacy policy based on the offers he receives from the third-parties who are interested in the database. We assume that the data collector responds independently to each data request from each third-party for each purpose. With this assumption, we only consider a single third-party with a single data usage purpose since any multi-third-party/multi-purpose case can be described as an aggregation of single third-party/single purpose independent cases. Note that if the same third-party requests the same data for two different purposes with different granularity levels, there is a potential of inferring additional information by combining the two versions of data. This situation can be modeled as two separate games and in the second game, the data has a higher economic value to the third-party. A more realistic and complicated approach would be modeling such cases with a single repeated game. Solving the problem using the latter approach is a future fork we are interested in.

We narrow our attention to a third-party with *aggregate reasoning* as the goal of data analysis. More specifically, we assume that the third-party only wants to run a COUNT-query on the data field df_j. The query retrieves the number of data records satisfying some conditions on df_j.

Finally, the data collector and the third-party are assumed to be rational and to know the probability function that describes data provider's privacy behavior. The model assumes that the the third-party knows the data collector's payoff function.

5 Model Description

Game theory provides a formal approach to model situations where a group of decision makers (players) have to choose optimum actions considering the mutual effects of other players' decisions. The main components of a game are the players and the set of actions available to each player. A player's strategy is a sequence of actions he chooses to maximize a payoff function. The payoff to each player depends on the decisions made by the player and the other players in the game. The stable outcomes of games are often predicted using the concept of Nash equilibrium. A specific play of the game is a Nash equilibrium if none of the players can increase their payoff by unilaterally deviating from their strategy [19].

The normal form and the extensive form (or sequential form) are two common alternatives to model games. Normal form games model situations where all players have to make decisions without knowing other players' decisions. Extensive form game models are capable of capturing a certain order for players' turns to move. The orderings are illustrated in a tree. A node in the game tree is a point where a specific player has to make a decision. Different choices by the player create different subtrees (subgames) of the node. Every possible sequence of actions from the root to the leaves represents a *terminal history* and any path from the root to an intermediate node is referred to as a *history* [19]. "Subgame-perfect Equilibrium" is usually used in extensive form games to predict the game's steady states. In a subgame perfect equilibrium all subgames of the original game are also in a Nash equilibrium [19].

The actions of a data collector, third-party, and data providers can influence the final decision about a privacy policy. Each of these three entities are trying to maximize their benefits. Therefore, we model our problem in a game theoretic framework. Since there exists a logical order on players' turns to play, the game is explained as an extensive form game with perfect information. The game's sub game perfect equilibria (or equilibria for short) illustrate the behavior of each party in their efforts to optimize their gains and suggest the stable outcomes of the game.

5.1 Players

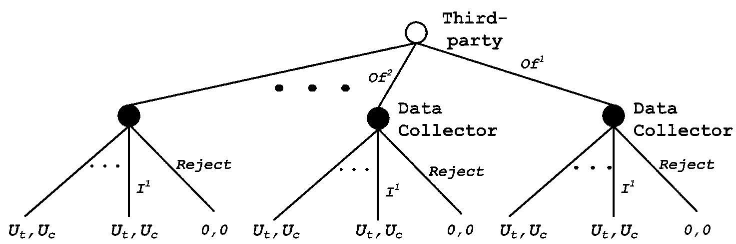

Fig. 1. Game tree without data providers' actions

The parties involved in the game are n potential data providers, a data collector c, and a third party t. As mentioned in Section 4, we assume a probabilistic function to explain data providers' mixed strategy at any point during the game. As a result, we only need to solve the game with regard to the data collector and third-party's decisions and take the actions of the data providers for granted.

The set of actions available to the third-party is making an offer, o, of the form $o = \langle g_1, ..., g_k, r, pr, p, Min, Max\rangle$.

In this offer g_i denotes the granularity level at which the third-party wishes to access data field df_i. Elements r, pr, and p represent retention period, purpose, and price respectively. Price is described as a real number and represents the amount of money the third-party pays for each data record at a specified privacy level. Finally, Min and Max parameters specify the minimum and maximum number of data records necessary for data analysis. In other words, if the database has less than Min number of records, the third-party is not interested in the database (and naturally does not pay) and by increasing the data records over the Max limit, the amount of money paid by the third-party for the database does not increase. In this paper, we assume that pr, Min, and Max are fixed for each instance of the problem.

We consider the situations where data analysis queries are of the following COUNT-query form:

```
SELECT COUNT(*) FROM privateTable WHERE Pred(df_j)
```

In this query `Pred` is a predicate defined on the data field df_j. Consequently, the payoff to the third-party, U_t, can be explained as follows:

$$U_t = \begin{cases} 0 & if\ size < Min \\ size(a.r - p) & if\ Min \leq size \leq Max \wedge g_j\ is\ odd \\ size(b.r - p) & if\ Min \leq size \leq Max \wedge g_j\ is\ even \\ Max(a.r - p) & if\ Max < size \quad \wedge g_j\ is\ odd \\ Max(b.r - p) & if\ Max < size \quad \wedge g_j\ is\ even \end{cases} \tag{5}$$

In this function $size$ is the number of data records provided to the third-party. Since the third-party only needs to work with data field df_j the payoff is not influenced by the granularity level of any other data field. Notice that if g_j is odd then the value of the data field df_j is partially revealed whereas even values for g_j imply exact revelation. In this sense, a and b are two parameters denoting how valuable a data item is if it is partial or exact. Therefore we have $a \leq b$. Values of a and b are predefined for each instance of the problem. They depend on the scale of generalization (or noise addition) process and how aligned it is with the COUNT-query predicate.

The data collector's actions are either rejecting the third-party's offer or attempting to collect information from data providers by choosing an incentive value I. The value of I is a real number and represents the monetary value of the incentive offered to data providers.

Let G be the cost of generalization or any other procedure to make the value partial and A represent the cost of anonymization. If the data collector decides to collect information according to an offer from the third-party then the cost of providing data field df_i at granularity level g_i (without considering the incentives) can be defined as:

$$CG(g_i) = \begin{cases} 0 & if\ g_i = 0\ \vee\ g_i = 2\ \vee\ g_i = 6 \\ G & if\ g_i = 1\ \vee\ g_i = 5 \\ A + G & if\ g_i = 3 \\ A & if\ g_i = 4 \end{cases} \tag{6}$$

Moreover, providing the database to the third-party is associated with a basic cost B for enforcing some data protection method to make sure that data practices of the third-party conform to the privacy policy.

Consider an offer o with the following description:

$o = \langle g_1, ..., g_k, r, pr, p, Min, Max \rangle$

We use C_o to denote the cost of providing the third-party with the database according to the offer o:

$$C_o = \sum_{i=1}^{k} CG(g_i) + B \tag{7}$$

The data collector also has to pay the promised incentive to the data providers. Therefore, the ultimate cost to the data collector, c, is:

$$Cost_c = \begin{cases} 0 & if\ c\ rejects \\ size * I + C_o & if\ c\ accepts \end{cases} \tag{8}$$

The benefit to the data collector, c, is:

$$Benefit_c = \begin{cases} 0 & if\ c\ rejects\ \vee size < Min \\ size * p & if\ c\ accepts\ \wedge Min \leq size \leq Max \\ Max * p & if\ c\ accepts\ \wedge Max \leq size \end{cases} \tag{9}$$

Consequently, the payoff to the data collector will be:

$$U_c = Benefit_c - Cost_c \tag{10}$$

By plugging the formulas of $Benefit_c$ and $Cost_c$ into Eq(10), we can formulate the utility to the data collector in the *acceptance case* as follows:

$$U_c^{accept} =$$

$$\begin{cases} 0 - size * I - C_o & if\ size < Min \\ size * (p - I) - C_o & if\ Min \leq size \leq Max \\ (Max)p - size * I - C_o & if\ Max < size \end{cases} \tag{11}$$

Table 1. The best strategies of the data collector in **Case1:** $n\alpha_o < Min$

	Condition	Best incentive	Best action†	Maximum payoff	Database size
a	$\frac{\gamma p-\alpha_o}{2\gamma} < \frac{Min-n\alpha_o}{n\gamma}$	$I = \frac{Min-n\alpha_o}{n\gamma}$	Accept if $U_c^{*,1a} > 0$ Reject if $U_c^{*,1a} < 0$ Indifferent otherwise	$\max\{U_c^{*,1a}, 0\}$	Min if Accept 0 if Reject
b	$\frac{Min-n\alpha_o}{n\gamma} \leq \frac{\gamma p-\alpha_o}{2\gamma} \leq \frac{Max-n\alpha_o}{n\gamma}$	$I = \frac{\gamma p-\alpha_o}{2\gamma}$	Accept if $U_c^{*,1b} > 0$ Reject if $U_c^{*,1b} < 0$ Indifferent otherwise	$\max\{U_c^{*,1b}, 0\}$	$n(\frac{\alpha_o+\gamma p}{2})$ if Accept 0 if Reject
c	$\frac{Max-n\alpha_o}{n\gamma} \leq \frac{\gamma p-\alpha_o}{2\gamma}$	$I = \frac{Max-n\alpha_o}{n\gamma}$	Accept if $U_c^{*,1c} > 0$ Reject if $U_c^{*,1c} < 0$ Indifferent otherwise	$\max\{U_c^{*,1c}, 0\}$	Max if Accept 0 if Reject

† Utilities are defined as: $U_c^{*,1a} = Min(p - \frac{Min-n\alpha_o}{n\gamma}) - C_o$, $U_c^{*,1b} = \frac{n}{\gamma}(\frac{\alpha_o+\gamma p}{2})^2 - C_o$, $U_c^{*,1c} = Max(p - \frac{Max-n\alpha_o}{n\gamma}) - C_o$.

5.2 Game Rules

The extensive-form game starts with a "one shot negotiation" between the third-party and the data collector. Then the privacy policy is established and data providers choose to participate or not. In practice, the negotiation can continue in multiple rounds but since we are not considering the time factor and the game is modeled with complete information, adding multiple rounds of negotiation does not change the results as long as it ends with an offer from the third-party followed by a response from the data collector. We believe that a negotiation started by a third-party (and not the data collector) is more natural and realistic. However, should the data collector initiate the negotiation, the analysis of the game and the outcome will potentially be different. Once the third-party makes an offer, the data collector can either reject the offer or collect the required information by publishing the privacy policy and announcing an incentive. In the former case, the payoff to both the third-party and the data collector are zero. In the latter case, the data providers have to decide whether to opt-in or opt-out. Generally, the data providers can get involved in a negotiation with the data collector to choose the best privacy level and incentive combination. However, modeling the game with multiple rounds of negotiation between the data collector and each data provider rapidly increases the depth of the game tree. Moreover, the data providers are usually not patient enough to go through such a negotiation for each piece of information they provide online. As explained in Section 4, considering each data provider's privacy preferences is practically infeasible. Therefore, the data providers' mixed strategy is explained with the probability model of Eq(3) and once the incentive and the parameters of the offer are plugged in, an exact probability will be given by the model for each case. Therefore, we can assume that optimal actions of data providers are already determined and the game is only illustrated to the point where the data collector makes a decision.

Except for the time when the data collector rejects, at each terminal node the size of the database is:

$$size = n * (\beta_0 + \beta_1 \frac{1}{1+g_1} + ... + \beta_k \frac{1}{1+g_k} + \theta \frac{1}{1+r} + \gamma I) \quad (12)$$

Where the values for the g_i's, r, and I are chosen by the third-party and data collector along the path that leads to the terminal node.

The payoffs to the third-party and the data collector are calculated by plugging the *size* into the payoff functions from Section 5.1. The (trimmed) game tree is depicted in Fig(1).

6 Data Collector's Best Responses

In the procedure of backward induction [19] on the trimmed tree, the first step is to find the optimal actions in the smallest subgames. These subgames are the ones that follow a history of the form ($o =$

$\langle g_1, ..., g_k, r, pr, p, Min, Max\rangle$) where o is an instance of an offer made by the third-party. In these subgames, it is the data collector's turn to either reject or accept (and announce some incentive) the third-party's offer.

After a history of the form o, the data collector optimizes his payoff based on the price p and other privacy attributes of the offer. At any of these nodes in the tree, the size of the database can be determined by Eq(12). To simplify the notations, let α_o denote the probability of opting in for a specific offer o with *zero incentive*. In other words:

$$\alpha_o = \beta_0 + \beta_1 \frac{1}{1+g_1} + ... + \beta_k \frac{1}{1+g_k} + \theta \frac{1}{1+r} \tag{13}$$

The values of g_i and r are plugged in from the specifications of o. Since we are assuming n potential data providers, the expected number of data providers who will opt-in with granularity g_i, retention r, and incentive I can be calculated as:

$$size = n(\alpha_0 + \gamma I) \tag{14}$$

When $size$ is plugged in into Eq(11) and the conditions are specified based on I, we can restate U_c^{accept} as follows:

$$U_c^{accept} =$$

$$\begin{cases} 0 - n(\alpha_o + \gamma I)I - C_o & if\ I < \frac{Min - n\alpha_0}{n\gamma} \\ n(\alpha_o + \gamma I)(p - I) - C_o & if\ \frac{Min - n\alpha_0}{n\gamma} \leq I \leq \frac{Max - n\alpha_0}{n\gamma} \\ (Max)p - n(\alpha_o + \gamma I)I - C_o & if\ \frac{Max - n\alpha_0}{n\gamma} < I \end{cases} \tag{15}$$

Since we can safely assume that $\alpha_o > 0$ the first case of the U_c^{accept} function always yields a negative payoff and therefore choosing an incentive $I < \frac{Min - n\alpha_o}{n\gamma}$ is always dominated by at least the reject action. To find the maximum of the second and the third pieces of the function we find the maximizing I for each piece of function by setting the derivative of the piece to zero. We then need to compare this maximizing value with the upper and lower bounds of I in the piece. There is also an implicit condition on valid values of I. This condition ensures that the maximizing incentives are not less than zero. To simplify these comparisons and organize the results, we consider three cases: case 1 happens when the lower bound of I in the second piece is greater than zero, case 2 happens when the lower bound is less than zero but the upper bound of I in the second piece is greater than zero, and case 3 occurs when the upper bound of I in the second piece is less than zero. Notice that these cases are specific to to each instance of the game and are not dependent on the offer received from the third party [1].

6.1 Case 1: $n\alpha_o < Min$

Table 2. The best strategies of the data collector in **Case 2:** $Min \leq n\alpha_o \leq Max$

	Condition	Best incentive	Best action†	Maximum payoff	Database size
a	$\frac{\gamma p - \alpha_o}{2\gamma} < 0$	$I = 0$	Accept if $U_c^{*,2a} > 0$ Reject if $U_c^{*,2a} < 0$ Indifferent otherwise	$\max\{U_c^{*,2a}, 0\}$	$n\alpha_o$ if Accept 0 if Reject
b	$0 \leq \frac{\gamma p - \alpha_o}{2\gamma} \leq \frac{Max - n\alpha_o}{n\gamma}$	$I = \frac{\gamma p - \alpha_o}{2\gamma}$	Accept if $U_c^{*,2b} > 0$ Reject if $U_c^{*,2b} < 0$ Indifferent otherwise	$\max\{U_c^{*,2b}, 0\}$	$n(\frac{\alpha_o + \gamma p}{2})$ if Accept 0 if Reject
c	$\frac{Max - n\alpha_o}{n\gamma} \leq \frac{\gamma p - \alpha_o}{2\gamma}$	$I = \frac{Max - n\alpha_o}{n\gamma}$	Accept if $U_c^{*,2c} > 0$ Reject if $U_c^{*,2c} < 0$ Indifferent otherwise	$\max\{U_c^{*,2c}, 0\}$	Max if Accept 0 if Reject

† Utilities are defined as: $U_c^{*,2a} = n\alpha_o p - C_o$, $U_c^{*,2b} = \frac{n}{\gamma}(\frac{\alpha_o + \gamma p}{2})^2 - C_o$, $U_c^{*,2c} = Max(p - \frac{Max - n\alpha_o}{n\gamma}) - C_o$.

Table 3. The best strategies of the data collector in **Case 3:** $Max < n\alpha_o$

Condition	Best incentive	Best action†	Maximum payoff	Database size
none	$I = 0$	Accept if $U_c^{*,3} > 0$ Reject if $U_c^{*,3} < 0$ Indifferent otherwise	$\max\{U_c^{*,3}, 0\}$	$n\alpha_o$ if Accept 0 if Reject

† Utility is defined as: $U_c^{*,3} = p(Max) - C_o$

This case happens when the anticipated number of data providers who opt-in for the policy *without any incentive* is less than Min.

If the maxima of either the second or the third piece of the U_c^{accept} function in Eq(15) is greater than zero then the data collector decides to accept and the number of data records can be calculated based on the maximizing incentive. If the maximum of these two pieces is

[1] These three cases cover the space of all possible values for upper and lower bounds on values of I in the second piece of Eq(15).

equal to zero then the data collector can either choose to accept or reject (both are optimal actions).

To find the maximum in the second piece of the function, we calculate the derivative of second piece of U_c^{accept} with respect to I and set it to zero:

$$\frac{dU_c^{accept}}{dI} = n\gamma p - n\alpha_o - 2n\gamma I = 0 \Rightarrow I^* = \frac{\gamma p - \alpha_o}{2\gamma} \tag{16}$$

Since the second derivative of the function is negative, the function is concave down and thus I^* represents a local maximum. Depending on the value of I^* the following three sub-cases can occur:

(a) If $I^* < \frac{Min - n\alpha_o}{n\gamma}$, then the maximum happens at the beginning of the interval ($i.e.$, $I = \frac{Min - n\alpha_o}{n\gamma}$). In this case the maximum payoff to the data collector (if he accepts), denoted by $U_c^{*,1a}$, is calculated as:

$$U_c^{*,1a} = Min(p - \frac{Min - n\alpha_o}{n\gamma}) - C_o \tag{17}$$

(b) If $\frac{Min - n\alpha_o}{n\gamma} \leq I^* \leq \frac{Max - n\alpha_o}{n\gamma}$, then the maximum happens at I^*. In this case the maximum payoff to the data collector (if he accepts), denoted by $U_c^{*,1b}$, is calculated as:

$$U_c^{*,1b} = \frac{n}{\gamma}(\frac{\alpha_o + \gamma p}{2})^2 - C_o \tag{18}$$

(c) If $\frac{Max - n\alpha_o}{n\gamma} \leq I^*$, then the maximum happens at the end of the interval ($i.e.$, $I = \frac{Max - n\alpha_o}{n\gamma}$). In this case the maximum payoff to the data collector (if he accepts), denoted by $U_c^{*,1c}$, is calculated as:

$$U_c^{*,1c} = Max(p - \frac{Max - n\alpha_o}{n\gamma}) - C_o \tag{19}$$

The maximum of the third piece of the function can be determined by finding the derivative of U_c^{accept} (third piece) with respect to I and setting it to zero:

$$\frac{dU_c^{accept}}{dI} = -n\alpha_o - 2n\gamma I = 0 \Rightarrow I^* = -\frac{\alpha_o}{2\gamma} < 0 < \frac{Max - n\alpha_o}{n\gamma} \tag{20}$$

Notice that since the second derivative is negative, the I^* is a local maximum. However, as Eq(20) shows, this I^* is less than the beginning of the interval and the maximizing incentive will be $I = \frac{Max - n\alpha_o}{n\gamma}$. Plugging this incentive in the third piece of the payoff function we receive the same maximum payoff as $U_c^{*,1c}$ in Eq(19). The best responses in case 1 are summarized in Table 1.

6.2 Case 2: $Min \le n\alpha_o \le Max$

This case happens when the anticipated number of data providers who opt-in for the policy *without any incentive* is more than (or equal to) Min but less than or equal to Max.

A procedure similar to Section 6.1 can be used to find the local optima of the payoff function. If the maxima of either the second or the third piece of the U_c^{accept} function in Eq(15) is greater than zero then the data collector decides to accept and the number of data records can be calculated based on the maximizing incentive. If the maximum of these two pieces is equal to zero then the data collector can either choose to accept or reject (both are optimal actions).

The maximum incentive in the second piece of the function, can be calculated by Eq(16). Based on the equation, $I^* = \frac{\gamma p - \alpha_o}{2\gamma}$ is the local maximum of the second piece of the function. Depending on the value of I^* the following three cases can occur:

(a) If $I^* < 0$, then the maximum happens at the zero. Notice that unlike case 1, the beginning of the interval is still less than zero and we cannot set I to the lower bound of this piece. In this case the maximum payoff to the data collector (if he accepts), denoted by $U_c^{*,2a}$, is calculated as:

$$U_c^{*,2a} = n\alpha_o p - C_o \tag{21}$$

(b) If $0 \le I^* \le \frac{Max - n\alpha_o}{n\gamma}$, then the maximum happens at I^*. In this case the maximum payoff to the data collector (if he accepts), denoted by $U_c^{*,1b}$, is calculated as:

$$U_c^{*,2b} = \frac{n}{\gamma}(\frac{\alpha_o + \gamma p}{2})^2 - C_o \tag{22}$$

(c) If $\frac{Max-n\alpha_o}{n\gamma} \leq I^*$, then the maximum happens at the end of the interval (*i.e.*, $I = \frac{Max-n\alpha_o}{n\gamma}$). In this case the maximum payoff to the data collector (if he accepts), denoted by $U_c^{*,2c}$, is calculated as:

$$U_c^{*,2c} = Max(p - \frac{Max - n\alpha_o}{n\gamma}) - C_o \tag{23}$$

Identical to case 1, the maximum of the third piece of the function can be determined by finding the derivative of U_c^{accept} (third piece) with respect to I and setting it to zero. Based on Eq(20), the local maximum is less than the beginning of the interval and therefore $I = \frac{Max-n\alpha_o}{n\gamma}$ maximizes the third piece of the function. This value for the incentive provides the same maximum payoff as $U_c^{*,2c}$ in Eq(23). The best responses in case 2 are summarized in Table 2.

6.3 Case 3: $Max < n\alpha_o$

In this last case, the number of data providers who are willing to share their information *without any incentive* is already greater than Max. Since more than Max opt-in with zero incentive, increasing the incentive above zero will entice more data providers without adding anything to the data collector's payoff. In this case, the best action of the data collector is to either accept with $I^* = 0$ or reject. With zero incentive, the payoff to the data collector would be $U_c^{*,3} = p(Max) - C_o$. This result is shown in Table 3.

7 Subgame Perfect Equilibria

To find the subgame perfect equilibria of the game, we need to take the best actions of the data collector from Section 6 as given and find the optimal actions of the third-party. The following three propositions help us to reduce the search space for the best actions.

Theorem 1. *Let* $o1 = \langle g_1, ..., g_j, ..., g_k, r, pr, p, Min, Max\rangle$ *be an offer such that at least one of the* g_i*'s with* $i \neq j$ *is set to a level higher than zero. Recall that* df_j *is the data field over which the predicate of the COUNT-query is defined. The third-party can do at least as good as* $o1$ *by making an offer* $o2 = \langle 0, 0, ..., g_j, ..., 0, r, pr, p, Min, Max\rangle$ *or* $o2' = \langle 0, 0, ..., g_j, ..., 0, r, pr, 0, Min, Max\rangle$*.*

Proof. Consider the description of α_o given in Eq(13). Since all parameters β_0,..., β_k are greater than zero and there is at least one g_i that is zero in $o2$ but more than zero in $o1$, we have $\alpha_{o1} < \alpha_{o2}$. Moreover, as one of the g_i 's in $o2$ changes from zero to another granularity level in $o1$, $CG(g_i)$ from Eq(6) either increases or stays the same. Thus, the inequality $C_{o2} \leq C_{o1}$ holds.

With these two facts we show that for all meaningful combinations of cases (from Section 6):

- **part 1 -** If the data collector accepts $o1$ he will also accept $o2$.
- **part 2 -** The expected database size after accepting offer $o2$ is at least as large as offer $o1$.
- **part 3 -** Offering $o2$ or $o2'$ provides the third-party with a payoff at least as large as $o1$.

Since $\alpha_{o1} < \alpha_{o2}$ not all combination of cases apply to offers $o1$ and $o2$. All possible combinations are enumerated in Table 4. For each combination, we first prove parts 1 and 2 and then justify part 3 of the theorem.

Table 4. Possible combinations of cases for $o1$ and $o2$

Case for $o2$ \ Case for $o1$	1a	1b	1c	2a	2b	2c	3
1a	✓	✗	✗	✗	✗	✗	✗
1b	✓	✓	✗	✗	✗	✗	✗
1c	✓	✓	✓	✗	✗	✗	✗
2a	✓	✓	✗	✓	✓	✗	✗
2b	✓	✓	✗	✗	✓	✗	✗
2c	✓	✓	✓	✗	✓	✓	✗
3	✓	✓	✓	✓	✓	✓	✓

Proof of parts 1 and 2 :

Case 1a for *o1* vs. case 1a for *o2* - If the data collector accepts offer $o1$ in case 1a, then the payoff $U_c^{*,1a}$ to the data collector for offer $o1$ is greater than or equal to zero[2]. In other words:

$$0 \leq Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} = U_c[o1]$$

Since $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, we have:

$$Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} \leq Min(p - \frac{Min - n\alpha_{o2}}{n\gamma}) - C_{o2} = U_c[o2]$$

Consequently, the data collector would also accept offer $o2$.

According to Table 1 if the data collector accepts any offer in case 1a the expected size of the database will be Min. Therefore, the expected database size after accepting offer $o2$ is at least as large as offer $o1$.

Case 1a for *o1* vs. case 1b for *o2* - If the data collector accepts offer $o1$ in case 1a, then the data collector's payoff $U_c^{*,1a}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} = U_c[o1]$$

Based on Eq(16), the maximum of the U_c^{accept} in case 1 happens by setting $I^* = \frac{\gamma p - \alpha_o}{2\gamma}$ and this maximum yields $U_c^{*,1b}$. Therefore, for every offer, $U_c^{*,1a} \leq U_c^{*,1b}$. We have:

$$Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} \leq \frac{n}{\gamma}(\frac{\alpha_{o1} + \gamma p}{2})^2 - C_{o1}$$

Since $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, we have:

$$0 \leq \frac{n}{\gamma}(\frac{\alpha_{o1}+\gamma p}{2})^2 - C_{o1} \leq \frac{n}{\gamma}(\frac{\alpha_{o2}+\gamma p}{2})^2 - C_{o2}$$
$$= U_c[o2]$$

Therefore, $o2$ will be accepted.

[2] In the rest of the proof, we denote the data collector's payoff for offer o as $U_c[o]$ and don't include the superscripts. The superscripts can be deduced based on the case that applies to the offer.

According to Table 1 if the data collector accepts an offer in case 1a the size of the database will be Min and if he accepts an offer in case 1b the expected size would be $n(\frac{\alpha_{o2}+\gamma p}{2})$. Since case 1b applies to offer $o2$, the condition for this case is fulfilled:

$$\frac{Min-n\alpha_{o2}}{n\gamma} \leq \frac{\gamma p-\alpha_{o2}}{2\gamma} \Rightarrow$$

$$\frac{Min}{n} \leq \frac{\gamma p}{2} + \frac{\alpha_{o2}}{2} \qquad \Rightarrow$$

$$Min \leq n(\frac{\gamma p+\alpha_{o2}}{2})$$

Therefore, the expected database size after accepting offer $o2$ is at least as large as offer $o1$.

Case 1a for *o1* vs. case 1c for *o2* - If the data collector accepts offer $o1$ in case 1a, then the data collector's payoff $U_c^{*,1a}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} = U_c[o1]$$

Since $Min \leq Max$ and $C_{o2} \leq C_{o1}$, we have:

$$Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} \leq Max(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o2}$$

We now show that $(p - \frac{Min-n\alpha_{o1}}{n\gamma}) \leq (p - \frac{Max-n\alpha_{o2}}{n\gamma})$ and therefore $0 \leq U_c[o1] \leq U_c[o2]$.

Since case 1c applies to $o2$, based on the condition of this case in Table 1 we have:

$$\frac{Max - n\alpha_{o2}}{n\gamma} \leq \frac{\gamma p - \alpha_{o2}}{2\gamma}$$

Since $\alpha_{o1} \leq \alpha_{o2}$, we also have:

$$\frac{\gamma p - \alpha_{o2}}{2\gamma} \leq \frac{\gamma p - \alpha_{o1}}{2\gamma}$$

Finally, condition 1a that applies to $o1$ requires:

$$\frac{\gamma p - \alpha_{o1}}{2\gamma} \leq \frac{Min - n\alpha_{o1}}{n\gamma}$$

Based on the past three inequalities, we have:

$$\frac{Max-n\alpha_{o2}}{n\gamma} \leq \frac{Min-n\alpha_{o1}}{n\gamma} \Rightarrow$$

$$-\frac{Min-n\alpha_{o1}}{n\gamma} \leq -\frac{Max-n\alpha_{o2}}{n\gamma} \Rightarrow$$

$$p-\frac{Min-n\alpha_{o1}}{n\gamma} \leq p-\frac{Max-n\alpha_{o2}}{n\gamma} \Rightarrow$$

$$Min(p-\frac{Min-n\alpha_{o1}}{n\gamma})-C_{o1} \leq Max(p-\frac{Max-n\alpha_{o2}}{n\gamma})-C_{o2} \Rightarrow$$

$$0 \leq U_c[o1] \leq U_c[o2]$$

Therefore, $o2$ will be accepted.

According to Table 1 if the data collector accepts an offer in case 1a the size of the database will be Min and if he accepts an offer in case 1c the expected size would be Max. Since $Min \leq Max$, the expected database size after accepting offer $o2$ is at least as large as offer $o1$.

Case 1a for *o1* vs. case 2a for *o2* - If the data collector accepts offer $o1$ in case 1a, then the data collector's payoff $U_c^{*,1a}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq Min(p-\frac{Min-n\alpha_{o1}}{n\gamma})-C_{o1}=U_c[o1]$$

Since $C_{o2} \leq C_{o1}$, we have:

$$0 \leq Min(p-\frac{Min-n\alpha_{o1}}{n\gamma})-C_{o1} \leq Min(p-\frac{Min-n\alpha_{o1}}{n\gamma})-C_{o2}$$

According to Table 1 the condition for case 1 is $n\alpha_{o1} < Min$. Therefore, by sunstituting the term α_{o1} with Min we get:

$$\begin{aligned} Min(p-\tfrac{Min-n\alpha_{o1}}{n\gamma})-C_{o2} &\leq Min(p-\tfrac{Min-Min}{n\gamma})-C_{o2} \\ &= Min\ p - C_{o2} \end{aligned}$$

Finally, case 2 applies to α_{o2} and the condition for case 2 (in Table 2) is $Min \leq n\alpha_{o2}$. With this property we substitute the term Min with $n\alpha_{o2}$ to get the following:

$$Min\ p - C_{o2} \leq n\alpha_{o2}p - C_{o2} = U_c[o2]$$

Based on the past four inequalities we conclude that the data collector's payoff $U_c^{*,2a}$ for offer $o2$ is greater than or equal to zero.

According to Table 1 if the data collector accepts an offer in case 1a the size of the database will be Min and if he accepts an offer in case 2a the expected size would be $n\alpha_{o2}$. The condition for case 2 (see Table 2) is $Min \leq n\alpha_{o2}$ and therefore, the expected database size after accepting offer $o2$ is at least as large as offer $o1$.

Case 1a for *o1* vs. case 2b for *o2* - If the data collector accepts offer $o1$ in case 1a, then he would also accept offer $o2$. The proof of this claim is identical to "case 1a for $o1$ vs. case 1b for $o2$".

According to Table 1 if the data collector accepts an offer in case 1a the size of the database will be Min and if he accepts in case 2b the expected size would be $n(\frac{\alpha_{o2}+\gamma p}{2})$. Case 2b applies to offer $o2$ and part of the condition for this case (see Table 2) is:

$$0 \leq \frac{\gamma p - \alpha_{o2}}{2\gamma} \Rightarrow$$

$$\alpha_{o2} \leq \gamma p \Rightarrow$$

$$n(\frac{\alpha_{o2}+\alpha_{o2}}{2}) \leq n(\frac{\alpha_{o2}+\gamma p}{2}) \Rightarrow$$

$$n\alpha_{o2} \leq n(\frac{\alpha_{o2}+\gamma p}{2})$$

Moreover, the condition for case 2 is $Min \leq n\alpha_{o2}$. Therefore, the expected database size after accepting offer $o2$ is at least as large as offer $o1$.

Case 1a for *o1* vs. case 2c for *o2* - If the data collector accepts offer $o1$ in case 1a, he would also accept the offer $o2$ in case 2c because $U_c^{*,2c} = U_c^{*,1c}$ and the proof is identical to "case 1a for $o1$ vs. case 1c for $o2$".

The expected size of the database after accepting offer $o2$ in case $2c$ is at least the same as the expected size of the database if $o1$ is accepted in case $1a$ since $Min \leq Max$.

Case 1a for *o1* vs. case 3 for *o2* - If the data collector accepts offer $o1$ in case 1a, then the data collector's payoff $U_c^{*,1a}$ for offer $o1$

is greater than or equal to zero. In other words:

$$0 \leq Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} = U_c[o1]$$

The condition for case 1 is $n\alpha_{o1} < Min$ (see Table 1). As a result:

$$0 \leq \frac{Min - n\alpha_{o1}}{n\gamma} \quad \Rightarrow$$

$$-\frac{Min - n\alpha_{o1}}{n\gamma} \leq 0$$

Consequently, the following holds:

$$Min(p - \frac{Min - n\alpha_{o1}}{n\gamma}) - C_{o1} \leq Min(p) - C_{o1}$$

Since $Min \leq Max$ and $C_{o2} \leq C_{o1}$, we have:

$$Min(p) - C_{o1} \leq Max(p) - C_{o2} = U_c[o2]$$

The equations prove that in this case offer $o2$ would also be accepted.

The expected size of the database after accepting offer $o1$ in case $1a$ is Min (see Table 2). The expected size of the database after accepting offer $o2$ in case 3 is $n\alpha_{o2}$ (see Table 3). Since $Min \leq Max \leq n\alpha_{o2}$ (the condition for case 3), if offer $o2$ is accepted the expected size is at least the same as offer $o1$.

Case 1b for *o1* vs. case 1b for *o2* - If the data collector accepts offer $o1$ in case 1b, then his payoff $U_c^{*,1b}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq \frac{n}{\gamma}(\frac{\alpha_{o1} + \gamma p}{2})^2 - C_{o1} = U_c[o1]$$

Since $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, we have:

$$\frac{n}{\gamma}(\frac{\alpha_{o1} + \gamma p}{2})^2 - C_{o1} \leq \frac{n}{\gamma}(\frac{\alpha_{o2} + \gamma p}{2})^2 - C_{o2} = U_c[o2]$$

Therefore, the data collector would also accept offer $o2$.

The expected size of the database if offer $o1$ is accepted would be $n(\frac{\alpha_{o1}+\gamma p}{2})$. The database size is $n(\frac{\alpha_{o2}+\gamma p}{2})$ if offer $o2$ is accepted. Since

$\alpha_{o1} \leq \alpha_{o2}$, if offer $o2$ is accepted the expected size of the dataset is at least as large as the expected size of the dataset after accepting offer $o1$.

Case 1b for *o1* vs. case 1c for *o2* - The data collector's payoff in case 1b can be rewritten as:

$$U_c^{*,1b} = n(\alpha_o + \gamma\frac{\gamma p - \alpha_o}{2\gamma})(p - \frac{\gamma p - \alpha_o}{2\gamma}) - C_o \qquad (24)$$

This equation is the result of plugging the optimum incentive $I^* = \frac{\gamma p - \alpha_o}{2\gamma}$ in the second piece of the U_c^{accept} function in Eq(11).

If the data collector accepts offer $o1$ in case 1b, his payoff $U_c^{*,1b}$ is greater than or equal to zero. In other words:

$$0 \leq n(\alpha_{o1} + \gamma\tfrac{\gamma p - \alpha_{o1}}{2\gamma})(p - \tfrac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} = U_c[o1] \Rightarrow$$

$$0 \leq n(\tfrac{\gamma p}{2} + \tfrac{\alpha_{o1}}{2})(p - \tfrac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} = U_c[o1]$$

According to Table 1, part of the condition for case 1b is:

$$\begin{aligned} &\tfrac{\gamma p - \alpha_{o1}}{2\gamma} \leq \tfrac{Max - n\alpha_{o1}}{n\gamma} \quad \Rightarrow \\ &\tfrac{\gamma p}{2} - \tfrac{\alpha_{o1}}{2} + \alpha_{o1} \leq \tfrac{Max}{n} \Rightarrow \qquad (25) \\ &n(\tfrac{\gamma p}{2} + \tfrac{\alpha_{o1}}{2}) \leq Max \end{aligned}$$

Based on this inequality we have:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} \leq Max(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1}$$

Since $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, the following holds:

$$Max(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} \leq Max(p - \frac{\gamma p - \alpha_{o2}}{2\gamma}) - C_{o2}$$

Based on the condition for case 1c in Table 1, we know:

$$\begin{aligned} &\tfrac{Max - n\alpha_{o2}}{n\gamma} \leq \tfrac{\gamma p - \alpha_{o2}}{2\gamma} \quad \Rightarrow \\ &-\tfrac{\gamma p - \alpha_{o2}}{2\gamma} \leq -\tfrac{Max - n\alpha_{o2}}{n\gamma} \end{aligned}$$

Consequently we have:

$$Max(p - \frac{\gamma p - \alpha_{o2}}{2\gamma}) - C_{o2} \leq Max(p - \frac{Max - n\alpha_{o2}}{n\gamma}) - C_{o2} = U_c[o2]$$

Therefore, the data collector would also accept offer $o2$ since the payoff would be greater than or equal to zero.

The expected size of the database is $n(\frac{\alpha_{o1}+\gamma p}{2})$ in case of accepting offer $o1$, and Max in case of accepting $o2$. Based on Eq(25), $n(\frac{\alpha_{o1}+\gamma p}{2}) \leq Max$ and accepting offer $o2$ would result in a database at least as large as the case where offer $o1$ is accepted.

Case 1b for *o1* vs. case 2a for *o2* - If the data collector accepts offer $o1$ in case 1b, his payoff $U_c^{*,1b}$ (see the version in Eq(24)) is greater than or equal to zero. In other words:

$$0 \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} = U_c[o1]$$

The condition for case 1 is $n\alpha_{o1} < Min$. Combining this fact with the condition in case 1b (see Table 1), proves the following inequality:

$$0 \leq \frac{Min - n\alpha_{o1}}{n\gamma} \leq \frac{\gamma p - \alpha_{o1}}{2\gamma}$$

As a result we have:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1}$$

Case 2a for $o2$ (see Table 2) implies that $\gamma p < \alpha_{o2}$. We also know that $\alpha_{o1} < \alpha_{o2}$ and $C_{o2} \leq C_{o1}$. Therefore we have:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1} \leq n(\frac{\alpha_{o2}}{2} + \frac{\alpha_{o2}}{2})p - C_{o2} = U_c[o2] \qquad (26)$$

The expected size of the database after accepting offer $o1$ would be $n(\frac{\alpha_{o1}+\gamma p}{2})$, and after accepting offer $o2$ would be $n\alpha_{o2}$. Since $\alpha_{o1} \leq \alpha_{o2}$ and $\gamma p < \alpha_{o2}$ (see condition 2a in Table 2), we have:

$$\begin{aligned} n(\tfrac{\alpha_{o1}+\gamma p}{2}) &\leq n(\tfrac{\alpha_{o2}+\gamma p}{2}) \\ &\leq n(\tfrac{\alpha_{o2}+\alpha_{o2}}{2}) \\ &= n\alpha_{o2} \end{aligned}$$

Therefore, accepting offer $o2$ would result in a database at least as large as the case where offer $o1$ is accepted.

Case 1b for *o1* vs. case 2b for *o2* - Since $U_c^{*,1b} = U_c^{*,2b}$ and the proof of "case 1b for $o1$ vs. case 1b for $o2$" only relies on the facts that $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, the proof of this case is identical to the proof of "case 1b for $o1$ vs. case 2b for $o2$".

Case 1b for *o1* vs. case 2c for *o2* - The proof of this case is identical to the proof of "case 1b for $o1$ vs. case 1c for $o2$". This is due to the facts that $U_c^{*,1c} = U_c^{*,2c}$, condition 1c (from Table 1) is the same as condition 2c (from Table 2), and the other inequalities used to prove "case 1b for $o1$ vs. case 1c for $o2$" are either $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, or related to $o1$.

Case 1b for *o1* vs. case 3 for *o2* - If the data collector accepts offer $o1$ in case 1b, his payoff $U_c^{*,1b}$ (see the version in Eq(24)) is greater than or equal to zero. In other words:

$$0 \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} = U_c[o1]$$

The condition for case 1 is $n\alpha_{o1} < Min$. Combining this fact with the condition in case 1b (see Table 1), proves the following inequality:

$$0 \leq \frac{Min - n\alpha_{o1}}{n\gamma} \leq \frac{\gamma p - \alpha_{o1}}{2\gamma}$$

As a result we have:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1}$$

Since $C_{o2} \leq C_{o1}$ and $n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2}) \leq Max$ (see Eq(25)), the following holds:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1} \leq Max\ p - C_{o2} = U_c[o2]$$

This proves that if offer $o1$ is accepted by the data collector, offer $o2$ is also accepted.

The expected size of the database after accepting offer $o1$ would be $n(\frac{\alpha_{o1}+\gamma p}{2})$ and after accepting offer $o2$ would be $n\alpha_{o2}$. Based on

Eq(25) we know $n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2}) \leq Max$. The condition for case 3 (see Table 3) is $Max \leq n\alpha_{o2}$. These two facts prove the following:

$$n(\frac{\alpha_{o1} + \gamma p}{2}) \leq Max \leq n\alpha_{o2}$$

Therefore, accepting offer $o2$ would result in a database at least as large as the case where offer $o1$ is accepted.

Case 1c for *o1* vs. case 1c for *o2* - If the data collector accepts offer $o1$ in case 1c, then his payoff $U_c^{*,1c}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq Max(p - \frac{Max - n\alpha_{o1}}{n\gamma}) - C_{o1} = U_c[o1]$$

Since $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, we have:

$$Max(p - \tfrac{Max - n\alpha_{o1}}{n\gamma}) - C_{o1} \leq Max(p - \tfrac{Max - n\alpha_{o2}}{n\gamma}) - C_{o2}$$
$$= U_c[o2]$$

Therefore, the data collector would also accept offer $o2$.

If any of the offers $o1$ or $o2$ are accepted the expected size of the database would be Max.

Case 1c for *o1* vs. case 2c for *o2* - Since the condition 1c (from Table 1) is the same as the condition 2c (from Table 2) and $U_c^{*,1c} = U_c^{*,2c}$, the proof of this case is identical to "case 1c for $o1$ vs. case 1c for $o2$".

Case 1c for *o1* vs. case 3 for *o2* - If the data collector accepts offer $o1$ in case 1c, then his payoff $U_c^{*,1c}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq Max(p - \frac{Max - n\alpha_{o1}}{n\gamma}) - C_{o1} = U_c[o1]$$

Based on the condition for case 1 (see Table 1), we know:

$$n\alpha_{o1} < Min < Max \Rightarrow$$

$$0 < Max - n\alpha_{o1} \quad \Rightarrow$$

$$-\tfrac{Max - n\alpha_{o1}}{n\gamma} < 0$$

Considering this inequality and the fact that $C_{o2} \leq C_{o1}$, we have:

$$Max(p - \tfrac{Max - n\alpha_{o1}}{n\gamma}) - C_{o1} \leq Max\ (p) - C_{o1} \qquad \Rightarrow$$

$$\leq Max\ (p) - C_{o2} = U_c[o2]$$

This proves that if offer $o1$ is accepted by the data collector, offer $o2$ is also accepted.

The expected size of the database after accepting offer $o1$ would be Max and after accepting offer $o2$ would be $n\alpha_{o2}$. Based on the condition for case 3 (see Table 3), we know $Max \leq n\alpha_{o2}$. Therefore, accepting offer $o2$ would result in a database at least as large as the case where offer $o1$ is accepted.

Case 2a for *o1* vs. case 2a for *o2* - If the data collector accepts offer $o1$ in case 2a then his payoff $U_c^{*,2a}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq n\alpha_{o1}p - C_{o1} = U_c[o1]$$

Since $\alpha_{o1} \leq \alpha_{o2}$ and $C_{o2} \leq C_{o1}$, we have:

$$n\alpha_{o1}p - C_{o1} \leq n\alpha_{o2}p - C_{o2} = U_c[o2]$$

Therefore, the data collector would also accept offer $o2$.

The expected size of the database would be $n\alpha_{o1}$ if offer $o1$ is accepted and $n\alpha_{o2}$ if offer $o2$ is accepted . Since $\alpha_{o1} \leq \alpha_{o2}$, the expected size of the database after accepting offer $o2$ is at least as large as accepting offer $o1$.

Case 2a for *o1* vs. case 3 for *o2* - If the data collector accepts offer $o1$ in case 2a then his payoff $U_c^{*,2a}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq n\alpha_{o1}p - C_{o1} = U_c[o1]$$

The condition for case 2 (see Table 2) is $n\alpha_{o1} \leq Max$. This inequality and the fact that $C_{o2} \leq C_{o1}$, prove the following:

$$n\alpha_{o1}p - C_{o1} \leq Max\ p - C_{o1} \leq Max\ p - C_{o2} = U_c[o2]$$

Therefore, the data collector would also accept offer $o2$.

The expected size of the database would be $n\alpha_{o1}$ if offer $o1$ is accepted and $n\alpha_{o2}$ if offer $o2$ is accepted . Since $\alpha_{o1} \leq \alpha_{o2}$, the expected size of the database after accepting offer $o2$ is at least as large as accepting offer $o1$.

Case 2b for *o1* vs. case 2a for *o2* - If the data collector accepts offer $o1$ in case 2b, his payoff $U_c^{*,2b}$ (see the version in Eq(24)) is greater than or equal to zero. In other words:

$$0 \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} = U_c[o1]$$

The condition for case 2b is $0 \leq \frac{\gamma p - \alpha_{o1}}{2\gamma}$. Therefore, the following inequality holds:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1}$$

Case 2a for $o2$ (see Table 2) implies that $\gamma p < \alpha_{o2}$. We also know that $\alpha_{o1} < \alpha_{o2}$ and $C_{o2} \leq C_{o1}$. Therefore we have:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1} \leq n(\frac{\alpha_{o2}}{2} + \frac{\alpha_{o2}}{2})p - C_{o2} = U_c[o2] \qquad (27)$$

The expected size of the database after accepting offer $o1$ would be $n(\frac{\alpha_{o1}+\gamma p}{2})$, and after accepting offer $o2$ would be $n\alpha_{o2}$. Since $\alpha_{o1} \leq \alpha_{o2}$ and $\gamma p < \alpha_{o2}$ (see condition 2a in Table 2), we have:

$$\begin{aligned} n(\tfrac{\alpha_{o1}+\gamma p}{2}) &\leq n(\tfrac{\alpha_{o2}+\gamma p}{2}) \\ &\leq n(\tfrac{\alpha_{o2}+\alpha_{o2}}{2}) \\ &= n\alpha_{o2} \end{aligned}$$

Therefore, accepting offer $o2$ would result in a dataset at least as large as the case where offer $o1$ is accepted.

Case 2b for *o1* vs. case 2b for *o2* - Since $U_c^{*,1b} = U_c^{*,2b}$, the proof of this case is identical to the proof of "case 2b for $o1$ vs. case 2b for $o2$".

Case 2b for *o1* vs. case 2c for *o2* - The proof of this case is identical to the proof of "case 1b for $o1$ and case 1c for $o2$". This is due to the facts that $U_c^{*,1b} = U_c^{*,2b}$, $U_c^{*,1c} = U_c^{*,2c}$, and conditions used to prove "case 1b for $o1$ vs. case 1c for $o2$" also hold in "case 2b for $o1$ vs. case 2c for $o2$".

Case 2b for *o1* vs. case 3 for *o2* - If the data collector accepts offer $o1$ in case 2b, his payoff $U_c^{*,2b}$ is greater than or equal to zero. In other words:

$$0 \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} = U_c[o1]$$

According to Table 2, the condition for case 2b is:

$$0 \leq \frac{\gamma p - \alpha_{o1}}{2\gamma}$$

As a result we have:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})(p - \frac{\gamma p - \alpha_{o1}}{2\gamma}) - C_{o1} \leq n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1}$$

Since $C_{o2} \leq C_{o1}$ and $n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2}) \leq Max$ (see Eq(25)), the following holds:

$$n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2})p - C_{o1} \leq Max\ p - C_{o2} = U_c[o2]$$

This proves that if offer $o1$ is accepted by the data collector, offer $o2$ is also accepted.

The expected size of the database after accepting offer $o1$ would be $n(\frac{\alpha_{o1}+\gamma p}{2})$ and after accepting offer $o2$ would be $n\alpha_{o2}$. Based on Eq(25) we know $n(\frac{\gamma p}{2} + \frac{\alpha_{o1}}{2}) \leq Max$. The condition for case 3 (see Table 3) is $Max \leq n\alpha_{o2}$. These two facts prove the following:

$$n(\frac{\alpha_{o1} + \gamma p}{2}) \leq Max \leq n\alpha_{o2}$$

Therefore, accepting offer $o2$ would result in a database at least as large as the case where offer $o1$ is accepted.

Case 2c for *o1* vs. case 2c for *o2* - Since the condition 1c (from Table 1) is the same as the condition 2c (from Table 2) and $U_c^{*,1c} = U_c^{*,2c}$, the proof of this case is identical to "case 1c for $o1$ vs. case 1c for $o2$".

Case 2c for *o1* vs. case 3 for *o2* - The proof of this case is identical to the proof of "case 1c for $o1$ and case 3 for $o2$". This is due to the facts that $U_c^{*,1c} = U_c^{*,2c}$, and all of the conditions used to prove "case 1c for $o1$ vs. case 3 for $o2$" also hold in "case 2c for $o1$ vs. case 3 for $o2$".

Case 3 for *o1* vs. case 3 for *o2* - If the data collector accepts offer $o1$ in case 3 then his payoff $U_c^{*,3}$ for offer $o1$ is greater than or equal to zero. In other words:

$$0 \leq p(Max) - C_{o1} = U_c[o1]$$

Since $C_{o2} \leq C_{o1}$, we have:

$$p(Max) - C_{o1} \leq p(Max) - C_{o2} = U_c[o2]$$

Therefore, the data collector would also accept offer $o2$.

The expected size of the database would be $n\alpha_{o1}$ if offer $o1$ is accepted and $n\alpha_{o2}$ if offer $o2$ is accepted . Since $\alpha_{o1} \leq \alpha_{o2}$, the expected size of the database after accepting offer $o2$ is at least as large as accepting offer $o1$.

Proof of part 3 :

So far we have shown that for every possible combination of cases that apply to offers $o1$ and $o2$, if the data collector accepts offer $o1$, he will also accept offer $o2$. Moreover, the data collector would expect a database of at lease the same size if he accepts offer $o2$ instead of offer $o1$. In this part we prove that offering $o2$ or $o2'$ (an offer with price 0) provides the third-party with a payoff at least as large as the payoff for offering $o1$.

In the rest of this proof we use ES_o to denote the expected size of the database if the data collector accepts offer o. We prove this part in the following three possible scenarios:

Scenario 1: Both offers get accepted by the data collector - In this case the payoff to the data collector for both offers $o1$ and $o2$ is greater than or equal to zero. Therefore, we anticipate that both offers get accepted. The expected size of the database would be ES_{o1} for offer $o1$ and ES_{o2} for offer $o2$. In part 2 of the proof we saw that $ES_{o1} \leq ES_{o2}$.

The payoff to the third-party is determined via Eq(5). In this function, if the required data field, df_j, is requested at a *partial* level (*i.e.*, g_j is odd) then the economic value of each of the the records is a to the third-party, and if the required data field, df_j, is requested at an *exact* level (*i.e.*, g_j is even) then the economic value of each of the the collected records is b to third-party. Therefore we have $a \leq b$. Without loss of generality, we only prove this scenario with the assumption that g_j is odd (*i.e.*, df_j is requested at the partial level in both $o1$ and $o2$). To prove the case where g_j is even, we have to substitute all occurrences of parameter a with parameter b in the proof. With the assumption "g_j is odd" the U_t function from Eq(5) becomes the following:

$$U_t[o] = \begin{cases} 0 & if\ ES_o < Min \\ ES_o(a.r - p) & if\ Min \leq ES_o \leq Max \\ Max(a.r - p) & if\ Max < ES_o \end{cases} \tag{28}$$

Notice the use of ES_o instead of the term $size$ in the original formula.

In all cases where offer $o1$ gets accepted (and consequently offer $o2$ gets accepted), ES_{o1} (and ES_{o2}) is at least Min. Therefore the first piece of the function in Eq(28) will never happen in this scenario. The following cases can happen:

- $ES_{o1} \leq Max$ and $ES_{o2} \leq Max$- For this case we have:

$$U_t[o1] = ES_{o1}(a.r - p)$$

$$U_t[o2] = ES_{o2}(a.r - p)$$

Since $ES_{o1} \leq ES_{o2}$, we conclude that $U_t[o1] \leq U_t[o2]$.

– $ES_{o1} \leq Max$ and $Max \leq ES_{o2}$- For this case we have:

$$U_t[o1] = ES_{o1}(a.r - p)$$

$$U_t[o2] = Max(a.r - p)$$

Since $ES_{o1} \leq Max$, we conclude that $U_t[o1] \leq U_t[o2]$.

– $Max \leq ES_{o1}$ and $Max \leq ES_{o2}$- For this case we have:

$$U_t[o1] = U_t[o2] = Max(a.r - p)$$

Therefore, we proved that if partial data is requested for df_j then for the scenario where both offers $o1$ and $o2$ are acceptable, the third-party makes at least as much profit by offering $o2$ instead of $o1$.

Scenario 2: Offer $o1$ does not get accepted but offer $o2$ gets accepted by the data collector - In this scenario, $U_c[o1] < 0$ and therefore we anticipate that the data collector rejects offer $o1$. When offer $o1$ is rejected, the payoff to the third-party is $U_t[o1] = 0$. However, in this scenario the third-party expects the data collector to accept offer $o2$ since $U_c[o2] \geq 0$. After plugging ES_{o2} for $size$ in Eq(5), if $U_t[o2] \geq 0$ then the third-party can make more profit by offering $o2$ instead of $o1$. If $U_t[o2] < 0$, then the third-party can make offer $o2'$ (ask for information without paying for it or simply not making an offer) and have a guaranteed payoff of at least zero. Therefore, we proved that in this scenario the third-party can make at least as much profit if he offers $o2$ or $o2'$ instead of $o1$.

Scenario 3: None of the offers get accepted by the data collector - In this scenario, $U_c[o1] < 0$ and $U_c[o2] < 0$. Therefore, we anticipate the data collector reject both of the offers. When the offers are rejected, the payoff to the third-party is zero. Consequently, the third-party can make at least as much profit by offering $o2$ instead of $o1$.

Theorem 2. *Let $o1 = \langle 0, 0, ..., g_j, ..., 0, r, pr, p, Min, Max\rangle$ be an offer such that $g_j \in \{3, 5\}$ where df_j is the data field over which the predicate of COUNT-query is defined. The third-party can do at least as good as $o1$ by making an offer $o2 = \langle 0, 0, ..., 1, ..., 0, r, pr, p, Min, Max\rangle$.*

Table 5. Potentially optimal payoffs to the third-party if $g_j = 1$

	Maximum payoff[†]	Subject to
$p1$	$U_t^{*,p1} = Min(ar^* - p^*)$	**[c-1]:** $n\alpha_{op} < Min$ **[c-1a]:** $\alpha_{op} + \gamma p < \frac{2Min}{n}$ **[c-p1]:** $\frac{\gamma C_{op}}{Min} + \frac{Min}{n} \leq \alpha_{op} + \gamma p$
$p2$	$U_t^{*,p2} = \frac{n}{2}(\alpha_{op} + \gamma p)(ar^* - p^*)$	**[c-1]:** $n\alpha_{op} < Min$ **[c-1b]:** $\frac{2Min}{n} \leq \alpha_{op} + \gamma p \leq \frac{2Max}{n}$ **[c-p2]:** $2\sqrt{\frac{\gamma C_{op}}{n}} \leq \alpha_{op} + \gamma p$
$p3$	$U_t^{*,p3} = Max(ar^* - p^*)$	**[c-1]:** $n\alpha_{op} < Min$ **[c-1c]:** $\frac{2Max}{n} \leq \alpha_{op} + \gamma p$ **[c-p3]:** $\frac{\gamma C_{op}}{Max} + \frac{Max}{n} \leq \alpha_{op} + \gamma p$
$p4$	$U_t^{*,p4} = n(\alpha_{op})(ar^* - p^*)$	**[c-2]:** $Min \leq n\alpha_{op} \leq Max$ **[c-2a]:** $\gamma p - \alpha_{op} < 0$ **[c-p4]:** $C_{op} \leq n\alpha_{op} p$
$p5$	$U_t^{*,p5} = \frac{n}{2}(\alpha_{op} + \gamma p)(ar^* - p^*)$	**[c-2]:** $Min \leq n\alpha_{op} \leq Max$ **[c-2b]:** $0 \leq \gamma p - \alpha_{op}$ and $\alpha_{op} + \gamma p \leq \frac{2Max}{n}$ **[c-p5]:** $2\sqrt{\frac{\gamma C_{op}}{n}} \leq \alpha_{op} + \gamma p$
$p6$	$U_t^{*,p6} = Max(ar^* - p^*)$	**[c-2]:** $Min \leq n\alpha_{op} \leq Max$ **[c-2c]:** $\frac{2Max}{n} \leq \alpha_{op} + \gamma p$ **[c-p6]:** $\frac{\gamma C_{op}}{Max} + \frac{Max}{n} \leq \alpha_{op} + \gamma p$
$p7$	$U_t^{*,p7} = Max(ar^* - p^*)$	**[c-3]:** $Max < n\alpha_{op}$ **[c-p7]:** $C_{op} \leq Max\ p$

[†] In all formulas r^* and p^* are the values that maximize the payoff in the row subject to the constraints. Parameters α_{op} and C_{op} are defined as: $\alpha_{op} = \beta_0 + \beta_1 + \ldots + \frac{\beta_j}{2} + \beta_{j+1} + \ldots + \beta_k + \frac{\theta}{r^*+1}$, $C_{op} = G + B$.

Proof. (Sketch) Consider the description of α_o given in Eq(13). Since all parameters β_0,..., β_k are greater than zero and $\frac{1}{1+1} > \frac{1}{1+3} > \frac{1}{1+5}$, we have $\alpha_{o1} < \alpha_{o2}$. Moreover, based on Eq(6) we have $CG(1) = CG(5) < CG(3)$. Therefore we can conclude that $C_{o2} \leq C_{o1}$. The rest of the proof is the same as the proof sketch in Theorem 1.

Theorem 3. *Let* $o1 = \langle 0, 0, ..., g_j, ..., 0, r, pr, p, Min, Max\rangle$ *be an offer such that* $g_j \in \{4, 6\}$ *where* df_j *is the data field over which the predicate of COUNT-query is defined. The third-party can do at least as good as* $o1$ *by making an offer* $o2 = \langle 0, 0, ..., 2, ..., 0, r, pr, p, Min, Max\rangle$.

Proof. (Sketch) Similar to Proposition 2, we can show that $\alpha_{o1} < \alpha_{o2}$ and $C_{o2} \leq C_{o1}$. The rest of the proof is the same as the proof sketch in Theorem 1.

According to the three propositions, we can safely narrow down our attention to "partial" and "exact" offers of the forms
$op = \langle 0, 0, ..., 1, ..., 0, r, pr, p, Min, Max\rangle$ and
$oe = \langle 0, 0, ..., 2, ..., 0, r, pr, p, Min, Max\rangle$.

To find the third-party's best strategy we consider the two offers of "partial" and "exact" granularity levels separately. For each of the two possible types of offers any of the cases mentioned in Section 6 can happen. The best strategies of the data collector determines an expected size (explained in the last column of Tables 1, 2, and 3) for the data table in each sub-case. Plugging these *size* elements into the corresponding piece of U_t function in Eq(5) and finding the maximizing combination of parameters r and p completes the procedure of finding the game's subgame perfect equilibria.

Tables 5 and 6 summarize the potentially optimal actions of the third-party if he chooses $g_j = 1$ (partial) and $g_j = 2$ (exact), respectively. The cases for partial offers are numbered as $p1$ to $p7$ (see Table 5) and the cases for the exact offers are numbered as $e1$ to $e7$ (see Table 6). Each of these cases exactly correspond to one of the seven sub-cases that could apply to an offer (*i.e.*, sub-cases 1a, 1b, 1c, 2a, 2b, 2c, and 3 from Tables 1, 2, and 3). Each row in Tables 5 and 6 lists a two-variable function to be maximized subject to three conditions (or two conditions in cases $p7$ and $e7$). The first condition in a row is the table condition (*i.e.*, the condition specified above either of the Tables 1, 2, and 3) and the second condition (if exists) refers to the case condition specified in the corresponding row in one of the three Tables 1, 2, or 3. The last condition in each row ensures that the proposed offer provides the data collector with a utility greater than or equal to zero (otherwise, the offer will not get accepted).

If the third-party sets $p = 0$ then his payoff would be 0. We specify this final case with $U_t^{*,0} = 0$.

The tables must be considered as a semi-lookup table; For any instance of the problem with specific values for β_0, β_1,.., β_k, θ, γ, Min, and Max the maximizing values, r^* and p^*, of each row can be easily calculated subject to the conditions specified. The row which yields maximum payoff to the third-party is the winning row. The values of r^*, p^*, and g_j in the winning row specify the parameters that the third-party will use in the offer in a subgame perfect equilibrium. The granularity levels of other data fields could be set to values greater than zero if and only if doing so yields the same payoff for the third-party.

Table 6. Potentially optimum payoffs to the third-party if $g_j = 2$

	Maximum payoff†	Subject to
e1	$U_t^{*,e1} = Min(br^* - p^*)$	**[c-1]:** $n\alpha_{oe} < Min$ **[c-1a]:** $\alpha_{oe} + \gamma p < \frac{2Min}{n}$ **[c-e1]:** $\frac{\gamma C_{oe}}{Min} + \frac{Min}{n} \leq \alpha_{oe} + \gamma p$
e2	$U_t^{*,e2} = \frac{n}{2}(\alpha_{oe} + \gamma p)(br^* - p^*)$	**[c-1]:** $n\alpha_{oe} < Min$ **[c-1b]:** $\frac{2Min}{n} \leq \alpha_{oe} + \gamma p \leq \frac{2Max}{n}$ **[c-e2]:** $2\sqrt{\frac{\gamma C_{oe}}{n}} \leq \alpha_{oe} + \gamma p$
e3	$U_t^{*,e3} = Max(br^* - p^*)$	**[c-1]:** $n\alpha_{oe} < Min$ **[c-1c]:** $\frac{2Max}{n} \leq \alpha_{oe} + \gamma p$ **[c-e3]:** $\frac{\gamma C_{oe}}{Max} + \frac{Max}{n} \leq \alpha_{oe} + \gamma p$
e4	$U_t^{*,e4} = n(\alpha_{oe})(br^* - p^*)$	**[c-2]:** $Min \leq n\alpha_{oe} \leq Max$ **[c-2a]:** $\gamma p - \alpha_{oe} < 0$ **[c-e4]:** $C_{oe} \leq n\alpha_{oe} p$
e5	$U_t^{*,e5} = \frac{n}{2}(\alpha_{oe} + \gamma p)(br^* - p^*)$	**[c-2]:** $Min \leq n\alpha_{oe} \leq Max$ **[c-2b]:** $0 \leq \gamma p - \alpha_{oe}$ and $\alpha_{oe} + \gamma p \leq \frac{2Max}{n}$ **[c-e5]:** $2\sqrt{\frac{\gamma C_{oe}}{n}} \leq \alpha_{oe} + \gamma p$
e6	$U_t^{*,e6} = Max(br^* - p^*)$	**[c-2]:** $Min \leq n\alpha_{oe} \leq Max$ **[c-2c]:** $\frac{2Max}{n} \leq \alpha_{oe} + \gamma p$ **[c-e6]:** $\frac{\gamma C_{oe}}{Max} + \frac{Max}{n} \leq \alpha_{oe} + \gamma p$
e7	$U_t^{*,e7} = Max(br^* - p^*)$	**[c-3]:** $Max < n\alpha_{oe}$ **[c-e7]:** $C_{oe} \leq Max\ p$

† In all formulas r^* and p^* are the values that maximize the payoff in the row subject to the constraints. Parameters α_{oe} and C_{oe} are defined as: $\alpha_{oe} = \beta_0 + \beta_1 + \ldots + \frac{\beta_j}{3} + \beta_{j+1} + \ldots + \beta_k + \frac{\theta}{r^*+1}$, $C_{oe} = B$.

8 Results in a Simplified Scenario

To move one step further from the 15 cases explained in Table (5), Table (6), and $U_t^{*,0}$, we make another simplifying assumption and consider the situation where the third-party requires the data field df_j for only one year. In other words, storing the database for more than one year does not add any advantage to the third-party. For this assumption we can rewrite the payoff function to the third-party as:

$$U_t = \begin{cases} 0 & if\ size < Min \quad \vee\ r < 1 \\ size(a-p) & if\ Min \leq size \leq Max \wedge g_j\ is\ odd \\ size(b-p) & if\ Min \leq size \leq Max \wedge g_j\ is\ even \\ Max(a-p) & if\ Max < size \quad \wedge\ g_j\ is\ odd \\ Max(b-p) & if\ Max < size \quad \wedge\ g_j\ is\ even \end{cases} \tag{29}$$

This utility function is the same as Eq(5) where all occurrences of r are substituted by 1. Similar to Theorems 2 and 3 it can be shown that in all of the Equilibrium offers if $p > 0$ then $r = 1$ and setting r to any value greater than one cannot provide the third-party with a higher payoff than $r = 1$. Therefore, it is enough to only analyze offers,op and oe, with the following format:
$op = \langle 0,\ 0, ...,\ 1, ...,\ 0,\ 1,\ pr,\ p,\ Min,\ Max\rangle$
$oe = \langle 0,\ 0, ...,\ 2, ...,\ 0,\ 1,\ pr,\ p,\ Min,\ Max\rangle$

With this assumption, we are basically moving from optimizing two-variable utility functions to single variable ones (r is not considered as a variable anymore). In this Section we show how to find the maximum of third-party's payoff function in each of the cases $p1$, $p2$, ..., $p7$ from Table 5. The same procedure can be followed to find the maximum of the utility functions for an exact offer (cases listed in Table 6). In all cases we assume that $r = 1$, $\alpha_{op} = \beta_0 + \beta_1 + ... + \frac{\beta_j}{2} + \beta_{j+1} + ... + \beta_k + \frac{\theta}{2}$, and $C_{op} = G + B$. The maximizing price in each case is denoted by $p^{*,\langle case\rangle}$ where $\langle case\rangle$ is the ID of the specific case (followed by sub-cases) that the maximizing price belongs to.

Best action in case *p1* - The constraints in this case can be summarized as:

- [c-1]: $n\alpha_{op} < Min$

– [c-1a]: $\alpha_{op} + \gamma p < \frac{2Min}{n}$
– [c-p1]: $\frac{\gamma C_{op}}{Min} + \frac{Min}{n} \leq \alpha_{op} + \gamma p$

Since the lower bound for $\alpha_{op}+\gamma p$ must be less than it's upper bound we need the following inequality to hold:

$$\begin{aligned} &\frac{\gamma C_{op}}{Min} + \frac{Min}{n} < \frac{2Min}{n} \Rightarrow \\ &n\gamma C_{op} < Min^2 \end{aligned} \tag{30}$$

Payoff to the third-party is the maximum of the $U_t^{p1} = Min(a - p)$ (the first row of Table 5 with r^* being substituted by 1). U_t^{p1} is a decreasing function of price and hence the minimum value for p maximizes the function. In other words:

$$p^{*,p1} = \frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \tag{31}$$

Notice that this value is greater than zero (since $n\alpha_{op} < Min$). If we plug this value in the definition of U_t^{p1}, we get the following payoff:

$$U_t^{*,p1} = Min(a - \frac{C_{op}}{Min} - \frac{Min}{n\gamma} + \frac{\alpha_{op}}{\gamma}) \tag{32}$$

As a result, the inequality

$$\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \tag{33}$$

guarantees that the optimum price, $p^{*,p1}$, provides the third-party with a utility of at least zero. Otherwise, the third-party is better off by setting the price equal to zero as in case $U_t^{*,0}$. The optimum price, $p^{*,p1}$ results in $U_c^{*,1a} = 0$. Therefore, the data collector will be indifferent between accepting and rejecting. In this work we only analyze those equilibria in which the data collector chooses to accept when he is indifferent between his choices. If $U_t^{*,p1}$ is the maximum among all other relevant U_t^* 's, then in the sub game perfect Equilibrium of the game, the third-party makes an offer *op* with price equal to $p^{*,p1}$ and the data collector sets the incentive to $\frac{Min - n\alpha_{op}}{n\gamma}$.

Best action in case *p2* - The constraints in this case can be summarized as:

- [c-1]: $n\alpha_{op} < Min$
- [c-1b]: $\frac{2Min}{n} \leq \alpha_{op} + \gamma p \leq \frac{2Max}{n}$
- [c-p2]: $2\sqrt{\frac{\gamma C_{op}}{n}} \leq \alpha_{op} + \gamma p$

Since both of the lower bounds for γp must be less than or equal to it's upper bound we need the following inequality to hold:

$$2\sqrt{\frac{\gamma C_{op}}{n}} \leq \frac{2Max}{n} \Rightarrow$$
$$n\gamma C_{op} \leq Max^2 \tag{34}$$

Payoff to the third-party is the maximum of the $U_t^{p2} = \frac{n}{2}(\alpha_{op} + \gamma p)(a - p)$ (the second row of Table 5 with r^* being substituted by 1). To find the maximum, we find the derivative of U_t^{p2} with respect to p and set it to zero:

$$\frac{d(U_t^{p2})}{dp} = n(\frac{\gamma}{2}(a-p) - \frac{\alpha_{op} + \gamma p}{2}) = 0 \Rightarrow p^{*,p2} = \frac{\gamma a - \alpha_{op}}{2\gamma} \tag{35}$$

$p^{*,p2}$ maximizes U_t^{p2} but we need to make sure it is within the boundaries. Two situations can happen:

(a) $n\gamma C_{op} < Min^2$: In this situation, $2\sqrt{\frac{\gamma C_{op}}{n}} < \frac{2Min}{n}$. The value of $p^{*,p2}$ is within boundaries if:

$$\frac{2Min}{n} \leq \gamma p^{*,p2} + \alpha_{op} \leq \frac{2Max}{n} \Rightarrow$$
$$\frac{2Min}{n} - \alpha_{op} \leq \frac{\gamma a - \alpha_{op}}{2} \leq \frac{2Max}{n} - \alpha_{op} \Rightarrow \tag{36}$$
$$\frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$$

I. If Eq(36) holds, $p^{*,p2(a)I} = p^{*,p2} = \frac{\gamma a - \alpha_{op}}{2\gamma}$ is the optimum price and we have:

$$U_t^{*,p2(a)I} = \frac{n}{8\gamma}(\alpha_{op} + \gamma a)^2 \tag{37}$$

In this case, the payoff to the data collector is:

$$U_c^{*,1b} = \frac{n}{16\gamma}(\alpha_{op} + \gamma a)^2 - C_{op} \tag{38}$$

II. if $a < \frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$, then the optimum price is the beginning of the boundary. Therefore, $p^{*,p2(a)II} = \frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$.
The payoff to the third-party would be:

$$U_t^{*,p2(a)II} = Min(a - \frac{2Min}{n\gamma} + \frac{\alpha_{op}}{\gamma}) \qquad (39)$$

To guarantee a payoff of at least zero to the third-party, the following condition must hold:

$$\frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \qquad (40)$$

If all of the constraints hold and $U_t^{*,p2(a)II}$ is the maximum among all other relevant U_t^* 's then in the equilibrium, the third-party makes offer *op* with $p = \frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ and the data collector accepts with $I = \frac{Min - n\alpha_{op}}{n\gamma}$. The payoff to the data collector will be:

$$U_c^{*,1b} = \frac{Min^2}{n\gamma} - C_{op} \qquad (41)$$

III. If $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$, then the maximizing price is $p^{*,p2(a)III} = \frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$. In this case the maximum payoff to the third-party is:

$$U_t^{*,p2(a)III} = Max(a - \frac{2Max}{n\gamma} + \frac{\alpha_{op}}{\gamma}) \qquad (42)$$

Notice that in this situation $U_t^{*,p2(a)III} > 0$. The payoff to the data collector is:

$$U_C^{*,1b} = \frac{Max^2}{n\gamma} - C_{op} \qquad (43)$$

If the settings of the problem is aligned with the conditions of this situation and $U_t^{*,p2(a)III}$ is greater than all other relevant U_t^* 's, then in the subgame perfect equilibrium the third-party makes an offer of the form *op* with $p = \frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ and the data collector accepts with an incentive $I = \frac{Max - n\alpha_{op}}{n\gamma}$.

(b) $Min^2 \leq n\gamma C_{op}$: In this situation, $\frac{2Min}{n} \leq 2\sqrt{\frac{\gamma C_{op}}{n}}$. The value of $p^{*,p2}$ is within boundaries if:

$$2\sqrt{\frac{\gamma C_{op}}{n}} \leq \gamma p^{*,p2} + \alpha_{op} \leq \frac{2Max}{n} \Rightarrow$$
$$2\sqrt{\frac{\gamma C_{op}}{n}} - \alpha_{op} \leq \frac{\gamma a - \alpha_{op}}{2} \leq \frac{2Max}{n} - \alpha_{op} \Rightarrow \tag{44}$$
$$4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$$

I. If Eq(44) holds, $p^{*,p2(b)I} = p^{*,p2} = \frac{\gamma a - \alpha_{op}}{2\gamma}$ is the optimum price and the situation is the same as p2(a)I.

II. if $a < 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$, then the optimum price is the beginning of the boundary. Therefore, $p^{*,p2(b)II} = 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$. The payoff to the third-party would be:

$$U_t^{*,p2(b)II} = \sqrt{n\gamma C_{op}}(a - 2\sqrt{\frac{C_{op}}{n\gamma}} + \frac{\alpha_{op}}{\gamma}) \tag{45}$$

The third-party needs this payoff to be greater than or equal to zero in order to make such an offer. Therefore this case imposes another constraint:

$$2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \tag{46}$$

If all of the constraints hold and $U_t^{*,p2(b)II}$ is the maximum among all other relevant U_t^*'s then in the equilibrium, the third-party makes offer *op* with $p = 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ and the data collector accepts with $I = \sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$. The payoff to the data collector will be:

$$U_c^{*,1b} = \frac{n}{\gamma}(\gamma\sqrt{\frac{C_{op}}{n\gamma}})^2 - C_{op} = 0 \tag{47}$$

Here the data collector would be indifferent between accepting or rejecting. In this work we only analyze those equilibria in which the data collector chooses to accept.

III. If $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$, then the case is the same as p2(a)III.

Best action in case *p3* - The constraints in this case can be summarized as:

- [c-1]: $n\alpha_{op} < Min$
- [c-1c]: $\frac{2Max}{n} \leq \alpha_{op} + \gamma p$
- [c-p3]: $\frac{\gamma C_{op}}{Max} + \frac{Max}{n} \leq \alpha_{op} + \gamma p$

The payoff to the third-party is the maximum of $U_t^{p3} = Max(a - p)$ (the third row of Table 5 with r^* being substituted by 1). This function is maximized when p is minimized. Based on the relationship between the two lower bounds, the following two situations can happen:

(a) $Max^2 < n\gamma C_{op}$: In this situation, $\frac{2Max}{n} < \frac{\gamma C_{op}}{Max} + \frac{Max}{n}$. The minimum price p would be $p^{*,p3(a)} = \frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ and the maximum payoff would be:

$$U_t^{*,p3(a)} = Max(a - \frac{C_{op}}{Max} - \frac{Max}{n\gamma} + \frac{\alpha_{op}}{\gamma}) \tag{48}$$

The following condition must hold to guarantee a non-negative payoff to the third-party:

$$\frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \tag{49}$$

If the constraints hold and $U_t^{*,p3(a)}$ is the maximum among all other relevant U_t^* 's then in the equilibrium, the third-party makes offer op with $p = \frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$. The maximizing incentive after receiving such an offer is $I = \frac{Max - n\alpha_{op}}{n\gamma}$ and the payoff to the data collator would be:

$$U_c^{*,1c} = Max(\frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} - \frac{Max}{n\gamma} + \frac{\alpha_{op}}{\gamma}) - C_{op} = 0 \tag{50}$$

Here the data collector would be indifferent between accepting or rejecting but we only consider those equilibria in which the data collector chooses to accept.

(b) $n\gamma C_{op} \leq Max^2$: In this situation, $\frac{\gamma C_{op}}{Max} + \frac{Max}{n} \leq \frac{2Max}{n}$. The minimum price p would be $p^{*,p3(b)} = \frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ and the maximum payoff would be:.

$$U_t^{*,p3(b)} = Max(a - \frac{2Max}{n\gamma} + \frac{\alpha_{op}}{\gamma}) \tag{51}$$

With the following constraint, the third-party is guaranteed a non-negative payoff:

$$\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \tag{52}$$

If the constraints hold and $U_t^{*,p3(b)}$ is the maximum among all other relevant U_t^* 's then in the equilibrium, the third-party makes offer op with $p = \frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$. The maximizing incentive after receiving such an offer is $I = \frac{Max - n\alpha_{op}}{n\gamma}$ and the payoff to the data collator would be:

$$\begin{aligned} U_c^{*,1c} &= Max(\tfrac{2Max}{n\gamma} - \tfrac{\alpha_{op}}{\gamma} - \tfrac{Max}{n\gamma} + \tfrac{\alpha_{op}}{\gamma}) - C_{op} \\ &= \tfrac{Max^2}{n\gamma} - C_{op} > 0 \end{aligned} \tag{53}$$

Best action in case *p4* - The constraints in this case can be summarized as:

- [c-2]: $Min \leq n\alpha_{op} \leq Max$
- [c-2a]: $\gamma p - \alpha_{op} < 0$
- [c-p4]: $C_{op} \leq n\alpha_{op} p$

Since the lower bound for p must be less than or equal to it's upper bound, the following inequality must hold:

$$\begin{aligned} &\tfrac{C_{op}}{n\alpha_{op}} < \tfrac{\alpha_{op}}{\gamma} \qquad \Rightarrow \\ &n\gamma C_{op} < (n\alpha_{op})^2 \end{aligned} \tag{54}$$

The payoff to the third-party is the maximum of $U_t^{p4} = n\alpha_{op}(a - p)$ (the fourth row of Table 5 with r^* being substituted by 1). This function is maximized when p is minimized. According to condition [c-p4], the minimum price value is $p^{*,p4} = \frac{C_{op}}{n\alpha_{op}}$. Consequently, the maximum payoff to the third-party would be:

$$U_t^{*,p4} = n\alpha_{op}(a - \frac{C_{op}}{n\alpha_{op}}) = n\alpha_{op}a - C_{op} \tag{55}$$

The third-party does not make such an offer unless his payoff for it is non-negative. In other words:

$$\frac{C_{op}}{n\alpha_{op}} \leq a \tag{56}$$

If the constraints hold and $U_t^{*,p4}$ is the maximum among all other relevant U_t^* 's then in the equilibrium, the third-party makes offer op with $p = \frac{C_{op}}{n\alpha_{op}}$. The maximizing incentive after receiving such an offer is $I = 0$ and the payoff to the data collator would be:

$$U_c^{*,2a} = n\alpha_{op}\frac{C_{op}}{n\alpha_{op}} - C_{op} = 0 \tag{57}$$

Here the data collector would be indifferent between accepting or rejecting.

Best action in case *p5* - The constraints in this case can be summarized as:

- [c- 2]: $Min \leq n\alpha_{op} \leq Max$
- [c-1b]:$0 \leq \gamma p - \alpha_{op}$ and $\alpha_{op} + \gamma p \leq \frac{2Max}{n}$
- [c-p5]: $2\sqrt{\frac{\gamma C_{op}}{n}} \leq \alpha_{op} + \gamma p$

Since the both of the lower bounds for $\gamma p + \alpha_{op}$ must be less than or equal to it's upper bound we need the following inequality to hold:

$$\begin{aligned} &2\sqrt{\frac{\gamma C_{op}}{n}} \leq \frac{2Max}{n} \Rightarrow \\ &n\gamma C_{op} \leq Max^2 \end{aligned} \tag{58}$$

Payoff to the third-party is the maximum of the $U_t^{p5} = \frac{n}{2}(\alpha_{op} + \gamma p)(a - p)$ (the fifth row of Table 5 with r^* being substituted by 1).

To find the maximum, we find the derivative of U_t^{p5} with respect to p and set it to zero:

$$\frac{d(U_t^{p5})}{dp} = n(\frac{\gamma}{2}(a-p) - \frac{\alpha_{op}+\gamma p}{2}) = 0 \Rightarrow p^{*,p5} = \frac{\gamma a - \alpha_{op}}{2\gamma} \quad (59)$$

$p^{*,p5}$ maximizes U_t^{p5} but we need to make sure it is within the boundaries. Two situations can happen:

(a) $(n\alpha_{op})^2 \leq n\gamma C_{op}$: In this situation, $\frac{\alpha_{op}}{\gamma} \leq \frac{2\sqrt{\frac{\gamma C_{op}}{n}}-\alpha_{op}}{\gamma}$. The value of $p^{*,p5}$ is within boundaries if:

$$\begin{aligned}
&2\sqrt{\frac{\gamma C_{op}}{n}} \leq \gamma p^{*,p5} + \alpha_{op} \leq \frac{2Max}{n} &&\Rightarrow \\
&2\sqrt{\frac{\gamma C_{op}}{n}} \leq \frac{\gamma a}{2} + \frac{\alpha_{op}}{2} \leq \frac{2Max}{n} &&\Rightarrow \\
&4\sqrt{\frac{\gamma C_{op}}{n}} - \alpha_{op} \leq \gamma a \leq \frac{4Max}{n} - \alpha_{op} &&\Rightarrow \\
&4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}
\end{aligned} \quad (60)$$

I. If Eq(60) holds, $p^{*,p5(a)I} = p^{*,p5} = \frac{\gamma a - \alpha_{op}}{2\gamma}$ is the optimum price and we have:

$$U_t^{*,p5(a)I} = \frac{n}{8\gamma}(\alpha_{op}+\gamma a)^2 \quad (61)$$

In this case, the payoff to the data collector is:

$$U_c^{*,2b} = \frac{n}{16\gamma}(\alpha_{op}+\gamma a)^2 - C_{op} \quad (62)$$

II. if $a < 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$, then the optimum price is the beginning of the boundary. Therefore, $p^{*,p5(a)II} = 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$. The payoff to the third-party would be:

$$\begin{aligned} U_t^{*,p5(a)II} &= \tfrac{n}{2}(\alpha_{op} + \gamma p^{*,p5(a)II})(a - p^{*,p5(a)II}) \\ &= \tfrac{n}{2}(\alpha_{op} + 2\sqrt{\tfrac{\gamma C_{op}}{n}} - \alpha_{op})(a - 2\sqrt{\tfrac{C_{op}}{n\gamma}} + \tfrac{\alpha_{op}}{\gamma}) \\ &= \sqrt{nC_{op}\gamma}(a - 2\sqrt{\tfrac{C_{op}}{n\gamma}} + \tfrac{\alpha_{op}}{\gamma}) \end{aligned} \tag{63}$$

A payoff of at least zero can be guaranteed for the third-party if the following holds:

$$2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \tag{64}$$

If all of the constraints hold and $U_t^{*,p5(a)II}$ is the maximum among all other relevant U_t^* 's then in the equilibrium, the third-party makes offer *op* with $p = 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ and the data collector accepts with $I = \sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$. The payoff to the data collector will be:

$$\begin{aligned} U_c^{*,2b} &= \tfrac{n}{\gamma}(\tfrac{\alpha_{op}+\gamma p^{*,p5(a)II}}{2})^2 - C_{op} \\ &= \tfrac{n}{\gamma}(\tfrac{\alpha_{op}+2\sqrt{\frac{\gamma C_{op}}{n}}-\alpha_{op}}{2})^2 = 0 \end{aligned} \tag{65}$$

Here the data collector would be indifferent between accepting or rejecting. We only consider those equilibria in which the data collector accepts.

III. If $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$, then the maximizing price is $p^{*,p5(a)III} = \frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$. In this case the maximum payoff to the third-party is:

$$\begin{aligned} U_t^{*,p5(a)III} &= \tfrac{n}{2}(\alpha_{op} + \gamma p^{*,p5})(a - p^{*,p5(a)III}) \\ &= Max(a - \tfrac{2Max}{n\gamma} + \tfrac{\alpha_{op}}{\gamma}) \end{aligned} \tag{66}$$

Notice that in this situation $U_t^{*,p5(a)III} > 0$. The payoff to the data collector is:

$$
\begin{aligned}
U_C^{*,2b} &= \frac{n}{\gamma}\left(\frac{\alpha_{op}+\gamma p^{*,p5(a)III}}{2}\right)^2 - C_{op} \\
&= \frac{Max^2}{n\gamma} - C_{op}
\end{aligned} \tag{67}
$$

If the settings of the problem is aligned with the conditions of this situation and $U_t^{*,p5(a)III}$ is greater than all other relevant U_t^*'s, then in the subgame perfect equilibrium the third-party makes an offer of the form *op* with price $p = \frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ and the data collector accepts with an incentive $I = \frac{Max - n\alpha_{op}}{n\gamma}$.

(b) $n\gamma C_{op} < (n\alpha_{op})^2$: In this situation, $\frac{2\sqrt{\frac{\gamma C_{op}}{n}} - \alpha_{op}}{\gamma} < \frac{\alpha_{op}}{\gamma}$. The value of $p^{*,p5}$ is within boundaries if:

$$
\begin{aligned}
&\alpha_{op} \le \gamma p^{*,p5} \le \frac{2Max}{n} - \alpha_{op} \;\Rightarrow \\
&\alpha_{op} \le \frac{\gamma a - \alpha_{op}}{2} \le \frac{2Max}{n} - \alpha_{op} \Rightarrow \\
&3\frac{\alpha_{op}}{\gamma} \le a \le \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}
\end{aligned} \tag{68}
$$

I. If Eq(68) holds, $P^{*,p5(b)I} = p^{*,p5} = \frac{\gamma a - \alpha_{op}}{2\gamma}$ is the optimum price and the situation is the same as P5(a)I.

II. if $a < 3\frac{\alpha_{op}}{\gamma}$, then the optimum price is the beginning of the boundary. Therefore, $p^{*,p5(b)II} = \frac{\alpha_{op}}{\gamma}$. The payoff to the third-party would be:

$$
U_t^{*,p5(b)II} = n\alpha_{op}\left(a - \frac{\alpha_{op}}{\gamma}\right) \tag{69}
$$

The third-party does not make this offer unless his payoff is at least zero. Therefore, the following constraint must hold:

$$
\frac{\alpha_{op}}{\gamma} \le a \tag{70}
$$

If all of the constraints hold and $U_t^{*,p5(b)II}$ is the maximum among all other relevant U_t^*'s then in the equilibrium, the third-party makes offer *op* with $p = \frac{\alpha_{op}}{\gamma}$ and the data collector accepts with $I = \frac{\alpha_{op} - \alpha_{op}}{2\gamma} = 0$. The payoff to the data collector will be:

$$\begin{aligned} U_c^{*,2b} &= \frac{n}{\gamma}(\frac{\alpha_{op} + \alpha_{op}}{2})^2 - C_{op} \\ &= \frac{n\alpha_{op}^2}{\gamma} - C_{op} > 0 \end{aligned} \quad (71)$$

III. If $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$, then the case is the same as p5(a)-III.

Best action in case *p6* - The constraints in this case can be summarized as:

- [c-2]: $Min \leq n\alpha_{op} \leq Max$
- [c-2c]: $\frac{2Max}{n} \leq \alpha_{op} + \gamma p$
- [c-p6]: $\frac{\gamma C_{op}}{Max} + \frac{Max}{n} \leq \alpha_{op} + \gamma p$

The payoff to the third-party is the maximum of $U_t^{p6} = Max(a - p)$ (the sixth row of Table 5 with r^* being substituted by 1). This function is maximized when p is minimized. Based on the relationship between the two lower bounds, the following two situations can happen:

(a) $Max^2 < n\gamma C_{op}$: The constraints and payoff functions in this sub-case are identical to case p3(a).
(b) $n\gamma C_{op} \leq Max^2$: The constraints and payoff functions in this sub-case are identical to case p3(b).

Best action in case *p7* - The constraints in this case can be summarized as:

- [c-3]: $Max < n\alpha_{op}$
- [c-p7]: $C_{op} \leq Max\ p$

The payoff to the third-party is the maximum of $U_t^{p7} = Max(a - p)$ (the seventh row of Table 5 with r^* being substituted by 1). This function is maximized when p is minimized. According to condition [c-p7], the minimum price value is $p^{*,p7} = \frac{C_{op}}{Max}$. Consequently, the maximum payoff to the third-party would be:

$$U_t^{*,p7} = Max(a - \frac{C_{op}}{Max}) \tag{72}$$

With the following constraint, the payoff to the third-party for such an offer would be non-negative:

$$\frac{C_{op}}{Max} \leq a \tag{73}$$

If the constraints hold and $U_t^{*,p7}$ is the maximum among all other relevant U_t^*'s then in the equilibrium, the third-party makes offer op with $p = \frac{C_{op}}{Max}$. The maximizing incentive after receiving such an offer is $I = 0$ and the payoff to the data collator would be:

$$U_c^{*,3} = Max\frac{C_{op}}{Max} - C_{op} = 0 \tag{74}$$

Here, the data collector would be indifferent between accepting or rejecting. We only study those equilibria in which he accepts.

We showed how to find the maximum of each utility function in Table 5. The procedure of finding maximum of utility functions in Table 6 is identical to the last seven cases (explained for partial granularity offers), if all occurrences of α_{op} are substituted by $\alpha_{oe} = \beta_0 + \beta_1 + ... + \frac{\beta_j}{3} + \beta_{j+1} + ... + \beta_k + \frac{\theta}{2}$ and all occurrences of C_{op} are substituted by $C_{oe} = B$.

8.1 Summary of the Simplified Scenario Results

We have already enumerated all possible cases that can apply to an offer (seven major cases and their sub-cases for each partial and exact granularity offers) and shown the maximum payoff to the players in each case. Some of the conditions in each case are independent of the maximizing price p^*. We call these conditions *environmental conditions* since they depend on the characteristics of an instance of the game (and are independent of a specific *play* of the game). Often, these conditions are shared between multiple cases. In this

Section we classify the cases based on the *environmental conditions* and in some classes find the case that yields the maximum payoff to the third-party (the equilibrium of the game) among all cases that apply to that class. We show how to classify the cases for an offer that asks for data field df_j at a partial granularity level (*i.e.*, offer op). The procedure is identical for an exact granularity offer oe. In the classes we have $\alpha_{op} = \beta_0 + \beta_1 + \ldots + \frac{\beta_j}{2} + \beta_{j+1} + \ldots + \beta_k + \frac{\theta}{2}$ and $C_{op} = G + B$.

Class 1 and subclasses- The environmental conditions in this class can be summarized as:

1. $n\alpha_{op} < Min$
2. $n\gamma C_{op} < Min^2$

The first condition in this class applies to cases $p1$, $p2$, and $p3$. Within these cases, the sub cases $p1$, $p2(a)$, and $p3(b)$ also require the second condition to be true (for case $p3(b)$ the second condition is a sufficient condition but not necessary).

The relevant cases can be further organized based on bounds on parameter a. Remember that parameter a denotes the real economic value that data item df_j has for the third-party (when the data item is access at a partial granularity level). Eq(33), Eq(36), Eq(40), and Eq(52) specify different boundaries on values of a. Notice that the lower bound $\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ (from Eq(33)) is smaller than the lower bound $\frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ (from Eq(40)) since the second condition of this class requires: $n\gamma C_{op} < Min^2$. The lower bound $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ (from Eq(52)) could be either less than or greater than $\frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ (from Eq(36)). We consider both possibilities when we go through the details of each subclass. When we put these lower/upper bounds of a into an order, we get the following subclasses:

I. $a < \frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$
II. $\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$
III. $\frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$
IV. $\frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$
V. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$

We determine the cases that apply to each subclass (based on Eq(33), Eq(36), Eq(40), and Eq(52)) and by finding the cases the provide higher payoff values to the third-party, we describe the subgame equilibria of each subclass:

I. $a < \frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
In this subclass, none of the Eq(33), Eq(36), Eq(40), Eq(52) and $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$ hold and consequently no offer provides the third-party with a payoff of at least zero. In this subclass, the subgame perfect equilibrium is the play in which the third-party is not making an offer (we denote such situation by an offer with price zero) or requesting the data field at the exact granularity level (if it provides the third-party with a payoff greater than zero).

II. $\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
The only case that applies to this subclass is $p1$ (see Eq(33)). The subgame perfect equilibrium of the game is the set of strategies mentioned in case $p1$ or requesting the data field at the exact granularity level (if it provides the third-party with a higher payoff).

III. $\frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
Based on Eq(33) and Eq(40), cases $p1$ and $p2(a)II$ apply to this subclass. Moreover, if $Max \leq 2Min$ then we have:

$$\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \tag{75}$$

and case $p3(b)$ may also apply to this subclass.
By looking at the details of the case $p2(a)$, we see that according to Eq(35) and Eq(36), if the conditions of this subclass hold then $U_t^{*,p2(a)II}$ is the local maximum of the $U_t^{p2} = \frac{n}{2}(\alpha_{op} + \gamma p)(a - p)$. This implies:

$$U_t^{*,p2(a)II} \geq U_t^{*,p2(a)III} \tag{76}$$

Moreover, based on Eq(42) and Eq(51), we have:

$$U_t^{*,p2(a)III} = U_t^{*,p3(b)} \tag{77}$$

Consequently, for this class case $U_t^{*,p2(a)II} \geq U_t^{*,p3(b)}$ and $p2(a)II$ is a more profitable choice for the third-party than $p3(b)$.
The third-party's payoff in case $p2(a)II$ never gets higher than his payoff in case $p1$. To see the reason, notice that the second condition of this class is $n\gamma C_{op} < Min^2$. This condition requires that $\frac{C_{op}}{Min} < \frac{Min}{n\gamma}$. Therefore, we have:

$$\begin{aligned}
&\frac{C_{op}}{Min} < \frac{Min}{n\gamma} && \Rightarrow \\
&\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} < \frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} && \Rightarrow \\
&-(\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}) > -(\frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}) && \Rightarrow \\
&Min(a - (\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma})) > \\
&\qquad\qquad Min(a - (\frac{2Min}{n\gamma} - \frac{\alpha_{op}}{\gamma})) && \Rightarrow \\
&U_t^{*,p1} > U_t^{*,p2(a)II}
\end{aligned} \tag{78}$$

As a result, case $p1$ is provides a higher profit to the third-party than $p2(a)II$ (for the details of the payoff functions please see Eq(32) and Eq(39)). To sum up, case $p1$ is the subgame perfect equilibrium of this subclass (unless requesting the data field at the exact granularity level provides the third-party with a higher payoff).

IV. $\frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
Based on Eq(33) and Eq(36), cases $p1$ and $p2(a)I$ apply to this subclass. Moreover, since $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ case $p3(b)$ may also apply to this subclass.
The third-party's payoff in case $p3(b)$ never gets higher than his payoff in case $p2(a)I$. To see the reason, notice that $U_t^{*,p3(b)}$ from Eq(51) is the same as $U_t^{*,p2(a)III}$ from Eq(42). Eq(35) proves that $U_t^{*,p2(a)I}$ from Eq(37) is the maximum of the function $U_t^{*,p2}$. Therefore, we know that:

$$U_t^{*,p2(a)I} \geq U_t^{*,p2(a)III} = U_t^{*,p3(b)} \tag{79}$$

Consequently, for this subclass, $p2(a)I$ is a more profitable choice than $p3(b)$.
In this subclass, between the cases $p1$, $p2(a)I$, and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

V. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$:
Based on Eq(33), Eq(36), and Eq(52), cases $p1$ and $p2(a)III$ and $p3(b)$ apply to this subclass. By comparing Eq(42) with Eq(51), and Eq(43) with Eq(53) we see that case $p2(a)III$ and $p3(b)$ offer the same maximum payoffs to the third-party and the data collector. Therefore, they represent the same strategy profile.
In this subclass, between the cases $p1$, $p2(a)III$ (or $p3(b)$), and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

Class 2 and subclasses- The environmental conditions in this class can be summarized as:

1. $n\alpha_{op} < Min$
2. $Min^2 \leq n\gamma C_{op} \leq Max^2$

The first condition in this class applies to cases $p1$, $p2$, and $p3$. The second condition does not conform to the required inequality of case $p1$ (see Eq(30)). Within the remaining cases $p2$ and $p3$, the sub cases $p2(b)$, and $p3(b)$ also have the second condition.

The relevant cases can be further organized based on bounds on parameter a. Eq(44), Eq(46), and Eq(52) specify different boundaries on value of a. The lower bound $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ (from Eq(52)) is always greater than (or equal to) the lower bound $2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ (from Eq(46)) since the second condition of this class implies $\sqrt{\frac{C_{op}}{n\gamma}} \leq \frac{Max}{n\gamma}$. But, $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ could be either less than or greater than $4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ (from Eq(44)). We consider both possibilities when we go through the details of each subclass. When we put these lower/upper bounds of a into an order, we get the following subclasses:

I. $a < 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$
II. $2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$
III. $4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$
IV. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$

In the following, for each of these subclasses, we discuss the applicable cases (based on Eq(44), Eq(46), and Eq(52)) and compare the amount of profit they provide to the third-party.

I. $a < 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$:
In this subclass, none of the Eq(44), (46), (52), and $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$ (the necessary condition for case $p2(b)III$) hold and consequently no offer provides the third-party with a payoff of at least zero. In this subclass, the subgame perfect equilibrium is the play in which the third-party is not making an offer (we denote such situation by an offer with price zero) or requesting the data field at the exact granularity level (if it provides the third-party with a payoff greater than zero).
II. $2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$:
According to Eq(46) and since $a \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ case $p2(b)II$ applies to this subclass. Moreover, if $Max^2 \leq 4n\gamma C_{op}$ then we have:

$$\begin{aligned}
&Max^2 \leq 4n\gamma C_{op} && \Rightarrow \\
&Max \leq 2\sqrt{n\gamma C_{op}} && \Rightarrow \\
&\frac{2Max}{n\gamma} \leq 4\sqrt{\frac{C_{op}}{n\gamma}} && \Rightarrow \\
&\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}
\end{aligned} \tag{80}$$

and case $p3(b)$ may also apply to this subclass.

The third-party's payoff in case $p3(b)$ never gets higher than his payoff in case $p2(b)II$. To see the reason, notice that $U_t^{*,p3(b)}$ from Eq(51) is the same as $U_t^{*,p2(b)III} = U_t^{*,p2(a)III}$ from Eq(42). When a is within the limits defined for this subclass, $U_t^{*,p2(b)II}$ (see Eq(45)) is the maximum of the $U_t^{*,p2}$ function. Therefore

$$U_t^{*,p2(b)II} \geq U_t^{*,p2(b)III} = U_t^{*,p3(b)} \tag{81}$$

and for this subclass, case $p2(b)II$ provides a higher payoff to the third-party than $p3(b)$.
In this subclass, between the case $p2(b)II$ and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

III. $4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
Based on Eq(44), case $p2(b)I$ apply to this subclass. Moreover, since $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ case $p3(b)$ may also apply to this subclass.
Third-party's payoff in case $p3(b)$ never gets higher than his payoff in case $p2(b)I$. To see the reason, notice that $U_t^{*,p3(b)}$ from Eq(51) is the same as $U_t^{*,p2(b)III} = U_t^{*,p2(a)III}$ from Eq(42). Eq(35) proves that $U_t^{*,p2(b)I} = U_t^{*,p2(a)I}$ from Eq(37) is the maximum of the function $U_t^{*,p2}$ (with the constraint $Min^2 \leq n\gamma C_{op}$). Therefore, we know that:

$$U_t^{*,p2(b)I} \geq U_t^{*,p2(b)III} = U_t^{*,p3(b)} \tag{82}$$

Consequently, for this subclass, case $p2(b)I$ is a more profitable choice for the third-party than $p3(b)$.
In this subclass, between the case $p2(b)I$ and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

IV. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$:
The subclass condition matches the condition for case $p2(b)III$. Moreover, since $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ case $p3(b)$ also applies to this subclass.

Cases $p2(b)III$ and $p2(a)III$ are identical. Moreover, by comparing Eq(42) with Eq(51), and Eq(43) with Eq(53) we see that cases $p2(a)III$ and $p3(b)$ offer the same maximum payoffs to the third-party and the data collector since the third-party offers the same price in both cases. Therefore, cases $p2(b)III$ and $p3(b)$ represent the same strategy profile.
In this subclass, between the case $p2(b)III$ (or $p3(b)$) and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

Class 3 and subclasses- The environmental conditions in this class can be summarized as:

1. $n\alpha_{op} < Min$
2. $Max^2 < n\gamma C_{op}$

The first condition in this class applies to cases $p1$, $p2$, and $p3$. The second condition does not conform to the required inequalities of case $p1$ and $p2$ (see Eq(30) and Eq(34)). Within the remaining case $p3$, the sub-case $p3(a)$ is the only one conforming to the second condition.

Based on the required condition of sub-case $p3(a)$ (see Eq(49)), we can distinguish two subclasses for this class:

I. $a < \frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
In this subclass, Eq(49) doesn't hold and consequently no offer provides the third-party with a payoff of at least zero. The subgame perfect equilibrium is the play in which the third-party is not making an offer (we denote such situations by an offer with price zero) or requesting the data field at the exact granularity level (if it provides the third-party with a payoff greater than zero).

II. $\frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a$:
This condition conforms to Eq(49). Between the case $p3(a)$ and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the

subgame perfect equilibrium of the subclass.

Class 4 and subclasses- The environmental conditions in this class can be summarized as:

1. $Min \leq n\alpha_{op} \leq Max$
2. $n\gamma C_{op} < (n\alpha_{op})^2 < Max^2$

The first condition in this class applies to cases $p4$, $p5$, and $p6$. Within these cases, the sub cases $p4$, $p5(b)$, and $p6(b)$ also require the second condition to be true (for case $p6(b)$ the second condition is a sufficient condition but not necessary).

The relevant cases can be further organized based on bounds on parameter a. Eq(56), Eq(68), Eq(70), and Eq(52) specify different boundaries on value of a. Notice that the lower bound $\frac{C_{op}}{n\alpha_{op}}$ (from Eq(56)) is smaller than the lower bound $\frac{\alpha_{op}}{\gamma}$ (from Eq(70)) since the second condition of this class requires: $\gamma C_{op} < n(\alpha_{op})^2$. The lower bound $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ (from Eq(52)) could be either less than or greater than $\frac{3\alpha_{op}}{\gamma}$ (from Eq(68)). We consider both possibilities when we go through the details of each subclass. When we put these lower/upper bounds of a into an order, we get the following subclasses:

I. $a < \frac{C_{op}}{n\alpha_{op}}$
II. $\frac{C_{op}}{n\alpha_{op}} \leq a \leq \frac{\alpha_{op}}{\gamma}$
III. $\frac{\alpha_{op}}{\gamma} \leq a \leq \frac{3\alpha_{op}}{\gamma}$
IV. $\frac{3\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$
V. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$

In the following, we describe the subgame perfect equilibrium of each subclass by finding the most profitable (to the third-party) cases among all applicable cases:

I. $a < \frac{C_{op}}{n\alpha_{op}}$:
In this subclass, none of the Eq(56), Eq(68), Eq(70), Eq(52) and $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$ hold and consequently no offer provides the third-party with a payoff of at least zero. In this subclass, the subgame perfect equilibrium is the play in which the third-party is not making an offer (we denote such situation by an

offer with price zero) or requesting the data field at the exact granularity level (if it provides the third-party with a payoff greater than zero).

II. $\frac{C_{op}}{n\alpha_{op}} \leq a \leq \frac{\alpha_{op}}{\gamma}$:
The only case that applies to this subclass is $p4$ (see Eq(56)). The subgame perfect equilibrium of the game is the set of strategies mentioned in case $p4$ or requesting the data field at the exact granularity level (if it provides the third-party with a higher payoff).

III. $\frac{\alpha_{op}}{\gamma} \leq a \leq \frac{3\alpha_{op}}{\gamma}$:
Based on Eq(56) and Eq(70), cases $p4$ and $p5(b)II$ apply to this subclass. Moreover, if $Max \leq 2(n\alpha_{op})$ then we have:

$$\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{3\alpha_{op}}{\gamma} \tag{83}$$

and case $p6(b)$ may also apply to this subclass.
By looking at the details of the case $p5(b)$, we see that according to Eq(59) and Eq(68), if the conditions of this subclass hold then $U_t^{*,p5(b)II}$ is the local maximum of the $U_t^{p5} = \frac{n}{2}(\alpha_{op} + \gamma p)(a - p)$. This implies:

$$U_t^{*,p5(b)II} \geq U_t^{*,p5(b)III} \tag{84}$$

Moreover, the third-party's payoffs in cases $p6(b)$, $p3(b)$, $p5(a)III$, and $p5(b)III$ are the same. We have:

$$U_t^{*,p5(b)III} = U_t^{*,p6(b)} \tag{85}$$

Consequently, for this subclass $U_t^{*,p5(b)II} \geq U_t^{*,p6(b)}$ and $p5(b)II$ is a better choice than $p6(b)$.
Third-party's payoff in case $p5(b)II$ never gets higher than his payoff in case $p4$. To see the reason, notice that the second condition of this class is $n\gamma C_{op} < (n\alpha_{op})^2$. This condition requires

that $C_{op} < \frac{n(\alpha_{op})^2}{\gamma}$. Therefore, we have:

$$\begin{aligned} & C_{op} < \frac{n(\alpha_{op})^2}{\gamma} && \Rightarrow \\ & -C_{op} > -\frac{n(\alpha_{op})^2}{\gamma} && \Rightarrow \\ & n\alpha_{op}a - C_{op} > n\alpha_{op}a - \frac{n(\alpha_{op})^2}{\gamma} && \Rightarrow \\ & U_t^{*,p4} > U_t^{*,p5(b)II} \end{aligned} \tag{86}$$

As a result the case $p5(b)II$ is dominated by case $p4$ (for the details of the payoff functions please see Eq(55) and Eq(69)). To sum up, case $p4$ is the subgame perfect equilibrium of this subclass (unless requesting the data field at the exact granularity level provides the third-party with a higher payoff).

IV. $\frac{3\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
Based on Eq(56) and Eq(68), cases $p4$ and $p2(b)I$ apply to this subclass. Moreover, since $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ case $p6(b)$ may also apply to this subclass.
The third-party's payoff in case $p6(b)$ never gets higher than his payoff in case $p4(b)I$. To see the reason, notice that the third-party's payoffs in cases $p6(b)$, $p3(b)$, $p5(a)III$, and $p5(b)III$ are the same. We have:

$$U_t^{*,p5(b)III} = U_t^{*,p6(b)} \tag{87}$$

Eq(59) proves that $U_t^{*,p5(b)I}$ (equal to $U_t^{*,p5(a)I}$ from Eq(61)) is the maximum of the function $U_t^{*,p5} = \frac{n}{2}(\alpha_{op} + \gamma p)(a - p)$. Therefore, we know that:

$$U_t^{*,p5(b)I} \geq U_t^{*,p5(b)III} = U_t^{*,p6(b)} \tag{88}$$

Consequently, for this subclass, case $p5(b)I$ is a more profitable choice than $p6(b)$ for the third-party.
In this subclass, between the cases $p4$, $p5(b)I$, and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

V. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$:
Based on Eq(56), Eq(68), and Eq(52), cases $p4$ and $p5(b)III$, and $p6(b)$ apply to this subclass.
The third-party offers the same price in cases $p6(b)$, $p3(b)$, $p5(a)III$, and $p5(b)III$ and his payoffs are the same in all of these four cases. Therefore, $p5(b)III$ and $p6(b)$ represent the same strategy profile.
In this subclass, between the cases $p4$, $p5(b)III$ (or $p6(b)$), and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

Class 5 and subclasses- The environmental conditions in this class can be summarized as:

1. $Min \leq n\alpha_{op} \leq Max$
2. $(n\alpha_{op})^2 \leq n\gamma C_{op} \leq Max^2$

The first condition in this class applies to cases $p4$, $p5$, and $p6$. The second condition does not conform to the required inequality of case $p4$ (see Eq(54)). Within the remaining cases $p5$ and $p6$, the sub cases $p5(a)$, and $p6(b)$ also have the second condition.

The relevant cases can be further organized based on bounds on parameter a. Eq(60), Eq(64), and Eq(52) specify different boundaries on values of a. The lower bound $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ (from Eq(52)) is greater than (or equal to) the lower bound $2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ (from Eq(64)) since the second condition of this class implies $\sqrt{\frac{C_{op}}{n\gamma}} \leq \frac{Max}{n\gamma}$. But, $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ could be either less than or greater than $4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ (from Eq(60)). We consider both possibilities when we go through the details of each subclass. When we put these lower/upper bounds of a into an order, we get the following subclasses:

I. $a < 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$

II. $2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$

III. $4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$
IV. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$

For each of these subclasses, we enumerate the applicable cases and compare them to find the subgame perfect equilibrium of each subclass:

I. $a < 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$:
In this subclass, none of the Eq(60), Eq(64), Eq(52), and $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$ (the necessary condition for case $p5(a)III$) hold and consequently no offer provides the third-party with a payoff of at least zero. In this subclass, the subgame perfect equilibrium is the play in which the third-party is not making an offer (we denote such situations by an offer with price zero) or requesting the data field at the exact granularity level (if it provides the third-party with a payoff greater than zero).

II. $2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$:
According to Eq(64) and since $a \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$ case $p5(a)II$ applies to this subclass. Moreover, if $Max^2 \leq 4n\gamma C_{op}$ then we have:

$$\begin{aligned} & Max^2 \leq 4n\gamma C_{op} && \Rightarrow \\ & Max \leq 2\sqrt{n\gamma C_{op}} && \Rightarrow \\ & \frac{2Max}{n\gamma} \leq 4\sqrt{\frac{C_{op}}{n\gamma}} && \Rightarrow \\ & \frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \end{aligned} \tag{89}$$

and case $p6(b)$ may also apply to this subclass.
The third-party's payoff in case $p6(b)$ never gets higher than his payoff in case $p5(a)II$. To see the reason, notice that the third-party's payoffs in cases $p6(b)$, $p3(b)$, and $p5(a)III$ are the same. We have:

$$U_t^{*,p5(a)III} = U_t^{*,p6(b)} \tag{90}$$

When a is within the limits defined for this subclass, $U_t^{*,p5(a)II}$ (see Eq(63)) is the maximum of the $U_t^{p5} = \frac{n}{2}(\alpha_{op} + \gamma p)(a - p)$

function.Therefore, we know that:

$$U_t^{*,p5(a)II} \geq U_t^{*,p5(a)III} = U_t^{*,p6(b)} \tag{91}$$

Consequently, for this subclass, case $p6(b)$ cannot offer a higher profit to the third-party compared to $p5(a)II$.
In this subclass, between the case $p5(a)II$ and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

III. $4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$:
Based on Eq(60), case $p5(a)I$ apply to this subclass. Moreover, since $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ case $p6(b)$ may also apply to this subclass.
Third-party's payoff in case $p6(b)$ never gets higher than his payoff in case $p5(a)I$. To see the reason, notice that the third-party's payoffs in cases $p6(b)$, $p3(b)$,and $p5(a)III$ are the same. We have:

$$U_t^{*,p5(a)III} = U_t^{*,p6(b)} \tag{92}$$

Eq(59) proves that $U_t^{*,p5(a)I}$ from Eq(61) is the maximum of function $U_t^{p5} = \frac{n}{2}(\alpha_{op} + \gamma p)(a - p)$. Therefore, we know that:

$$U_t^{*,p5(a)I} \geq U_t^{*,p5(a)III} = U_t^{*,p6(b)} \tag{93}$$

Consequently, for this subclass, case $p5(a)I$ is a better choice for the third-party than $p6(b)$.
In this subclass, between the case $p5(a)I$ and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

IV. $\frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} < a$:
The subclass condition matches the condition for case $p5(a)III$. Moreover, since $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ case $p6(b)$ also applies to this subclass.
Cases $p6(b)$ and $p3(b)$ provide the same payoffs to the third-party and data collector. Moreover, by comparing Eq(66) with Eq(51), and Eq(67) with Eq(53) we see that cases $p5(a)III$ and

$p3(b)$ offer the same maximum payoffs to the third-party and the data collector since the third-party offers the same price in both cases. Therefore, $p5(a)III$ and $p6(b)$ represent the same strategy profile.
In this subclass, between the case $p5(a)III$ (or $p6(b)$) and requesting the data item at the exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

Class 6 and subclasses- The environmental conditions in this class can be summarized as:

1. $Min \leq n\alpha_{op} \leq Max$
2. $Max^2 < n\gamma C_{op}$

The first condition in this class applies to cases $p4$, $p5$, and $p6$. The second condition does not conform to the required inequalities of case $p4$ and $p5$ (see Eq(54) and Eq(58)). Within the remaining case $p6$, the sub-case $p6(a)$ is the only one conforming to the second condition. Since the strategy in $p6(a)$ is the same as $p3(a)$, the subclasses and their analysis for this class are the same as Class 3.

Class 7 and subclasses- The environmental condition in this class can be summarized as:

- $Max < n\alpha_{op}$

This condition only matches with the case $p7$. Based on the required condition of case $p7$ (see Eq(73)), we can distinguish two subclasses for this class:

I. $a < \frac{C_{op}}{Max}$:
In this subclass, Eq(73) doesn't hold and consequently no offer provides the third-party with a payoff of at least zero. In this subclass, the subgame perfect equilibrium is the play in which the third-party is not making an offer or requesting the data field at the exact granularity level (if it provides the third-party with a payoff greater than zero).

II. $\frac{C_{op}}{Max} \leq a$:

This condition conforms to Eq(73). Between the case $p7$ and requesting the data item in exact granularity level the one that provides a higher payoff to the third-party determines the game's subgame perfect equilibrium.

Tables 7 to 13 summarize subgame perfect equilibria of the game in different situations if requesting the data field at the partial granularity level is more beneficial to the third-party than the exact one. Similar to Tables 7 to 13, seven more tables can be produced by changing every instance of parameter a to b and using $\alpha_{oe} = \beta_0 + \ \beta_1 + \ ... \ + \ \frac{\beta_j}{3} \ + \ \beta_{j+1} + \ ... \ + \ \beta_k + \ \frac{\theta}{2}$ and $C_{oe} = B$ instead of α_{op} and C_{op}. Consequently, each real instance of the problem matches one row from Tables 7 to 13 and one row in the other series of seven tables for offer o_e. In fact, these tables partition the problem space into 22×22 different spaces and demonstrate the outcome and subgame perfect equilibriums of each case.

Table 7. subgame perfect equilibria strategies for Class 1[†]

	Condition[‡]	Best price	U_t^*	U_c^*
I	$a < \frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_1$	0	0	reject
II/III	$l_1 \leq a < \frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_2$	$\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$	$Min(a - (\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}))$	$U_c^{*,1a} = 0$
IV	$l_2 \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_3$	$\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ OR $\frac{\gamma a - \alpha_{op}}{2\gamma}$	$\max\{Min(a - (\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma})),$ $\frac{n}{8\gamma}(\alpha_{op} + \gamma a)^2\}$	$U_c^{*,1a} = 0$ OR $U_c^{*,1b} = \frac{n}{16\gamma}(\alpha_{op} + \gamma a)^2 - C_{op}$
V	$l_3 < a$	$\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$ OR $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$	$\max\{Min(a - (\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma})),$ $Max(a - (\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}))\}$	$U_c^{*,1a} = 0$ OR $U_c^{*,1b} = \frac{Max^2}{n\gamma} - C_{op}$

[†] In this class, data df_j is requested at the partial granularity level, $n\alpha_{op} < Min$, and $n\gamma C_{op} < Min^2$

[‡] In all formulas $\alpha_{op} = \beta_0 + \ \beta_1 + \ ... \ + \ \frac{\beta_j}{2} \ + \ \beta_{j+1} + \ ... \ + \ \beta_k + \ \frac{\theta}{2}$ and $C_{op} = G + B$.

In the next Section we illustrate how these results can be used in two simple case studies and what the formulas imply.

9 Case Studies

We explain the usage and implications of the seven classes in Section 8.1 via two synthetic case studies.

Case study 1: The first case represents an abstract instance of the problem. Consider a situation where the third-party requires data

Table 8. subgame perfect equilibria strategies for Class 2†

	Condition‡	Best price	U_t^*	U_c^*
I	$a < 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} = l_1$	0	0	reject
II	$l_1 \leq a < 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} = l_2$	$2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$	$\sqrt{n\gamma C_{op}}(a - 2\sqrt{\frac{C_{op}}{n\gamma}} + \frac{\alpha_{op}}{\gamma})$	$U_c^{*,1b} = 0$
III	$l_2 \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_3$	$\frac{\gamma a - \alpha_{op}}{2\gamma}$	$\frac{n}{8\gamma}(\alpha_{op} + \gamma a)^2$	$U_c^{*,1b} = \frac{n}{16\gamma}(\alpha_{op} + \gamma a)^2 - C_{op}$
IV	$l_3 \leq a$	$\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$	$Max(a - (\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}))$	$U_c^{*,1b} = \frac{Max^2}{n\gamma} - C_{op}$

† In this class, data df_j is requested at the partial granularity level, $n\alpha_{op} < Min$, and $Min^2 \leq n\gamma C_{op} \leq Max^2$

‡ In all formulas $\alpha_{op} = \beta_0 + \beta_1 + \ldots + \frac{\beta_j}{2} + \beta_{j+1} + \ldots + \beta_k + \frac{\theta}{2}$ and $C_{op} = G + B$.

Table 9. subgame perfect equilibria strategies for Class 3†

	Condition‡	Best price	U_t^*	U_c^*
I	$a < \frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_1$	0	0	reject
II	$l_1 \leq a$	$\frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$	$Max(a - (\frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}))$	$U_c^{*,1c} = 0$

† In this class, data df_j is requested at the partial granularity level, $n\alpha_{op} < Min$, and $Max^2 < n\gamma C_{op}$

‡ In all formulas $\alpha_{op} = \beta_0 + \beta_1 + \ldots + \frac{\beta_j}{2} + \beta_{j+1} + \ldots + \beta_k + \frac{\theta}{2}$ and $C_{op} = G + B$.

Table 10. subgame perfect equilibria strategies for Class 4†

	Condition‡	Best price	U_t^*	U_c^*
I	$a < \frac{C_{op}}{n\alpha_{op}} = l_1$	0	0	reject
II/III	$l_1 \leq a < \frac{3\alpha_{op}}{\gamma} = l_2$	$\frac{C_{op}}{n\alpha_{op}}$	$n\alpha_{op}(a - (\frac{C_{op}}{n\alpha_{op}}))$	$U_c^{*,2a} = 0$
IV	$l_2 \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_3$	$\frac{C_{op}}{n\alpha_{op}}$ OR $\frac{\gamma a - \alpha_{op}}{2\gamma}$	$\max\{n\alpha_{op}(a - (\frac{C_{op}}{n\alpha_{op}}))$ $\frac{n}{8\gamma}(\alpha_{op} + \gamma a)^2\}$	$U_c^{*,2a} = 0$ OR $U_c^{*,2b} = \frac{n}{16\gamma}(\alpha_{op} + \gamma a)^2 - C_{op}$
V	$l_3 \leq a$	$\frac{C_{op}}{n\alpha_{op}}$ OR $\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$	$\max\{n\alpha_{op}(a - (\frac{C_{op}}{n\alpha_{op}}))$ $Max(a - (\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}))\}$	$U_c^{*,2a} = 0$ OR $U_c^{*,2b} = \frac{Max^2}{n\gamma} - C_{op}$

† In this class, data df_j is requested at the partial granularity level, $Min \leq n\alpha_{op} \leq Max$, and $n\gamma C_{op} < (n\alpha_{op})^2 < Max^2$

‡ In all formulas $\alpha_{op} = \beta_0 + \beta_1 + \ldots + \frac{\beta_j}{2} + \beta_{j+1} + \ldots + \beta_k + \frac{\theta}{2}$ and $C_{op} = G + B$.

Table 11. subgame perfect equilibria strategies for Class 5†

	Condition‡	Best price	U_t^*	U_c^*
I	$a < 2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} = l_1$	0	0	reject
II	$l_1 \leq a < 4\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma} = l_2$	$2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}$	$\sqrt{C_{op}\gamma n}(a - (2\sqrt{\frac{C_{op}}{n\gamma}} - \frac{\alpha_{op}}{\gamma}))$	$U_c^{*,2b} = 0$
III	$l_2 \leq a \leq \frac{4Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_3$	$\frac{\gamma a - \alpha_{op}}{2\gamma}$	$\frac{n}{8\gamma}(\alpha_{op} + \gamma a)^2$	$U_c^{*,2b} = \frac{n}{16\gamma}(\alpha_{op} - \gamma a)^2 - C_{op}$
IV	$l_3 \leq a$	$\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$	$Max(a - (\frac{2Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}))$	$U_c^{*,2b} = \frac{Max^2}{n\gamma} - C_{op}$

† In this class, data df_j is requested at the partial granularity level, $Min \leq n\alpha_{op} \leq Max$, and $(n\alpha_{op})^2 \leq n\gamma C_{op} \leq Max^2$

‡ In all formulas $\alpha_{op} = \beta_0 + \beta_1 + \ldots + \frac{\beta_j}{2} + \beta_{j+1} + \ldots + \beta_k + \frac{\theta}{2}$ and $C_{op} = G + B$.

Table 12. subgame perfect equilibria strategies for Class 6†

	Condition‡	Best price	U_t^*	U_c^*
I	$a < \frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma} = l_1$	0	0	reject
II	$l_1 \leq a$	$\frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}$	$Max(a - (\frac{C_{op}}{Max} + \frac{Max}{n\gamma} - \frac{\alpha_{op}}{\gamma}))$	$U_c^{*,2c} = 0$

† In this class, data df_j is requested at the partial granularity level, $Min \leq n\alpha_{op} \leq Max$, and $Max^2 < n\gamma C_{op}$

‡ In all formulas $\alpha_{op} = \beta_0 + \beta_1 + \ ... \ + \frac{\beta_j}{2} + \beta_{j+1} + \ ... \ + \beta_k + \frac{\theta}{2}$ and $C_{op} = G + B$.

Table 13. subgame perfect equilibria strategies for Class 7†

	Condition‡	Best price	U_t^*	U_c^*
I	$a < \frac{C_{op}}{Max} = l_1$	0	0	reject
II	$l_1 \leq a$	$\frac{C_{op}}{Max}$	$Max(a - (\frac{C_{op}}{Max}))$	$U_c^{*,3} = 0$

† In this class, data df_j is requested at the partial granularity level, $Max < n\alpha_{op}$.

‡ In all formulas $\alpha_{op} = \beta_0 + \beta_1 + \ ... \ + \frac{\beta_j}{2} + \beta_{j+1} + \ ... \ + \beta_k + \frac{\theta}{2}$ and $C_{op} = G + B$.

field df_j for only one year.Therefore, it is enough to only analyze offers, *op* and *oe*, with the following formats:
$op = \langle 0,\ 0, ...,\ 1, ...,\ 0,\ 1,\ pr,\ p,\ Min,\ Max \rangle$
$oe = \langle 0,\ 0, ...,\ 2, ...,\ 0,\ 1,\ pr,\ p,\ Min,\ Max \rangle$
For any offer of partial information request, *op*, let α_{op} denote the probability of a data provider providing data at partial granularity level ($g_j = 1$) with *zero incentive* and C_{op} represent the cost of providing the third-party with the database according to *op*. In other words, $\alpha_{op} = \beta_0 + \beta_1 + \ ... \ + \beta_j/2 + \beta_{j+1} + \ ... \ + \beta_k + \theta/2$, and $C_{op} = G + B$. Similar to α_{op} and C_{op} we use the notations $\alpha_{oe} = \beta_0 + \beta_1 + \ ... \ + \beta_j/3 + \beta_{j+1} + \ ... \ + \beta_k + \theta/2$ and $C_{oe} = B$ to represent the same concepts where the offer is made for exact granularity level, *oe*.

Let the abstract instance of the problem have the following setups:

Condition *op*(a): $n\alpha_{op} < Min$,
Condition *op*(b): $n\gamma C_{op} < Min^2$,
Condition *op*(c): $\frac{C_{op}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{op}}{\gamma} \leq a \leq \frac{4Min}{n\gamma} - \frac{\alpha_{op}}{\gamma}$,
Condition *oe*(a): $n\alpha_{oe} < Min$,
Condition *oe*(b): $n\gamma C_{oe} < Min^2$ and
Condition *oe*(c): $\frac{C_{oe}}{Min} + \frac{Min}{n\gamma} - \frac{\alpha_{oe}}{\gamma} \leq b \leq \frac{4Min}{n\gamma} - \frac{\alpha_{oe}}{\gamma}$

The conditions mentioned for *op* matches only with the second row of Table(7). In other words, the environmental conditions and boundaries on a match classes 1-II and 1-III. In both of these classes the game's subgame perfect equilibria are the set of strategies specified in case $p1$ from Section 8 or requesting data at the exact granularity level. As a result, by requesting data at the partial granularity level, the third-party expects the following maximum payoff:

$$U_t^{*,p} = U_t^{*,p1} = Min(a - \frac{C_{op}}{Min} - \frac{Min}{n\gamma} + \frac{\alpha_{op}}{\gamma}) \tag{94}$$

Similarly, by requesting data at the exact granularity level, the third-party expects the following maximum payoff:

$$U_t^{*,e} = U_t^{*,e1} = Min(b - \frac{C_{oe}}{Min} - \frac{Min}{n\gamma} + \frac{\alpha_{oe}}{\gamma}) \tag{95}$$

The third-party has to make a final decision between op and oe based on $U_t^{*,p1}$ and $U_t^{*,e1}$. The third-party chooses to ask for exact information if the following holds:

$$Min(a-(\frac{C_{op}}{Min}+\frac{Min}{n\gamma}-\frac{\alpha_{op}}{\gamma})) < Min(b-(\frac{C_{oe}}{Min}+\frac{Min}{n\gamma}-\frac{\alpha_{oe}}{\gamma})) \tag{96}$$

Considering the facts that $\alpha_{op} - \alpha_{oe} = \frac{B_j}{6}$ and $C_{op} - C_{oe} = G$, we can rewrite the inequality in Eq(96) as:

$$\frac{B_j}{6\gamma} - \frac{G}{Min} < b - a \tag{97}$$

Therefore, if Eq(97) holds, the third-party asks for exact information on df_j. In case of an equality the third-party would be indifferent between asking for exact or partial information. Finally, if $b - a < \frac{B_j}{6\gamma} - \frac{G}{Min}$ then the third-party is better off by asking for partial information. This analysis shows:

1. If sharing data field df_j has a high impact on the public's privacy decisions (as B_j increases), the firms are forced to collect only partial information rather than exact information.
2. As the influence of incentive becomes less important on the public's privacy decisions (as γ decreases), the only way to collect personal information is to protect individual's privacy by using generalization (or other perturbation methods) on data.

3. When the cost of generalization G increases, asking for data at the exact granularity level becomes the most profitable choice of the third-party.
4. By increasing the minimum number of required records, the third party has no choice other than providing privacy to the data providers and ask for data at the partial granularity level.
5. The inequality can be considered as a reference point to recognize sensitive attributes from non-sensitive ones. In other words, if B_j is high enough to violate Eq(97), it indicates that df_j is sensitive. Notice that sensitivity of an attribute also depends on the data application $(b - a)$.

Case study 2: As a more concrete case study, we consider a situation where a pharmacy goods manufacturer (a third-party) is planning to launch a few new production lines. To make the best decision on what kind of goods to produce, the manufacturer needs a database containing the pharmacy-related shopping habits of habitants in the area. To this end, the manufacturer decides to ask for such information from the largest supermarket (a data collector) in the city. The supermarket records the `date, total amount payable`, and `items` purchased when each customer pays for his basket. The supermarket can offer some discount and seek for the customer's permission to provide this information to the pharmacy goods manufacturer (for each shopping trip). The supermarket knows its clients' privacy behavior (possibly based on some past experience) and can model it with the following probability model:

$$\begin{aligned} prob(opt - in) = 0.05 + 0.02\tfrac{1}{g_1+1} + 0.05\tfrac{1}{g_2+1} + 0.2\tfrac{1}{g_3+1} \\ +0.04\tfrac{1}{r+1} + 0.05I \end{aligned} \quad (98)$$

Where g_1, g_2, and g_3 represent granularity levels of `date, total amount payable`, and `items` respectively.

The third-party is only interested in accessing `item` information for a year. This information can be provided at the exact granularity level or at the partial granularity level (after values are generalized to broader categories). Other parameters of this problem instance are summarized in Table 14.

The third-party must decide on the granularity level, g_3, of data field `item`, and the price for each piece of information. We calculate α_{op} (probability of opt-in for partial level with zero incentive)

Table 14. Parameter settings

	Parameter	Value		Parameter	Value
1	Max	$20,000$	2	Min	$8,000$
3	B	\$1,000	4	G	\$100
5	a	\$5	6	b	\$10
7	n	$30,000$			

and C_{op} (cost of providing partial information to the third-party) as follows:

$$\begin{aligned} \alpha_{op} &= \beta_0 + \beta_1 \tfrac{1}{g_1+1} + \beta_2 \tfrac{1}{g_2+1} + \beta_3 \tfrac{1}{g_3+1} + \theta \tfrac{1}{r+1} \\ &= 0.05 + 0.02 \tfrac{1}{0+1} + 0.05 \tfrac{1}{0+1} + 0.2 \tfrac{1}{1+1} + 0.04 \tfrac{1}{1+1} \\ &= 0.24 \\ C_{op} &= \ G + B = \ 1,100 \end{aligned} \tag{99}$$

Similarly, the probability of opt-in for the exact granularity level with zero incentive, α_{oe}, and the cost of providing such information, C_{oe}, are calculated as:

$$\begin{aligned} \alpha_{oe} &= \beta_0 + \beta_1 \tfrac{1}{g_1+1} + \beta_2 \tfrac{1}{g_2+1} + \beta_3 \tfrac{1}{g_3+1} + \theta \tfrac{1}{r+1} \\ &= 0.05 + 0.02 \tfrac{1}{0+1} + 0.05 \tfrac{1}{0+1} + 0.2 \tfrac{1}{2+1} + 0.04 \tfrac{1}{1+1} \\ &= 0.2 \\ C_{oe} &= \ B = \ 1,000 \end{aligned} \tag{100}$$

We can easily see that the settings of this problem conform to the conditions ***op*(a), (b), (c)** and ***oe*(a), (b), (c)**. Therefore we have:

$$\begin{aligned} U_t^{*,p1} &= Min(a - \tfrac{C_{op}}{Min} - \tfrac{Min}{n\gamma} + \tfrac{\alpha_{op}}{\gamma}) \\ &= 34633 \\ U_t^{*,e1} &= Min(b - \tfrac{C_{oe}}{Min} - \tfrac{Min}{n\gamma} + \tfrac{\alpha_{oe}}{\gamma}) \\ &= 69400 \end{aligned} \tag{101}$$

Since $U_t^{*,e1} > U_t^{*,p1}$, the payoff to the third-party would be higher if the exact granularity level is asked for. Therefore, in the stable state of the game the subgame perfect Equilibrium suggests the following payoffs and prices:

$$\begin{aligned} p^* &= \tfrac{C_{oe}}{Min} - \tfrac{Min}{n\gamma} + \tfrac{\alpha_{oe}}{\gamma} = 1.3250 \\ U_t^* &= Min(b - \tfrac{C_{oe}}{Min} - \tfrac{Min}{n\gamma} + \tfrac{\alpha_{oe}}{\gamma}) = 69400 \\ U_c^* &= 0 \end{aligned} \tag{102}$$

Notice that the equilibrium results in $U_c = 0$. Therefore the data collector will be indifferent between accepting and rejecting. We have only analyzed those equilibria in which the data collector accepts when he is indifferent. Although accepting the offer does not add any monetary value to the utility of the data collector, the opportunity of building a relationship with a possibly future supplier might convince the supermarket to collect information. The data collector sets the incentive to $\frac{Min - n\alpha_{oe}}{n\gamma} = 1.2$.

10 Conclusions and Future Work

This study set out to determine a balanced privacy policy settings for aggregate query applications. By adopting a game theoretic framework, the optimization problem is analyzed from the viewpoints of the third-party and the data collector while considering the effects of data providers' privacy preferences. Backward induction method is used to find the games' subgame perfect equilibria. The results narrow down the solution space to a choice between the fifteen cases mentioned in Section 7. Moreover, for each problem instance, some of the cases do not need any further analysis since the conditions cannot be satisfied. Based on these cases, a simplified scenario is adapted to mathematically show how to find the subgame perfect equilibria of the game. We demonstrated the application of this study using two case studies.

Our work provides directions on how to set a privacy policy to achieve maximum revenue while respecting data providers' privacy preferences. We showed how to model and tackle the challenge for a simple privacy policy language and COUNT-query data application. But our model is not limited to a specific language and any statement-based privacy policy language can be modeled in our game as long as data granularity is somehow expressed and preferences of the players can be defined accordingly. After solving the game with our sample policy language and data application, our results demonstrate how a shift in data providers' privacy behavior can change the price of information and expectations of data collectors. These results can also indicate a metric to assess the sensitivity of each piece of information.

We have already used this model to address the challenge of setting privacy parameter values in sanitization systems in a different work [2]. This work completes our previous research and shows the applicability of our model to the privacy policy declaration approach and proves that our model is not exclusive to a specific privacy protection mechanism. Since our game model is generic and independent of a specific privacy protection method, it can be potentially used for benchmarking purposes in order to compare different privacy protection approaches from variety of perspectives. This is the main direction of our future work. Further research might explore the effects of price discrimination in offering the incentives, the influences of introducing bargaining to the game, and the outcome of the game if the third-party has incomplete information. Moreover, experimental investigations are needed to estimate the parameters of the data providers' privacy behavior model.

References

1. A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 3(1):26–33, 2005.
2. R. K. Adl, M. Askari, K. Barker, and R. Safavi-Naini. Privacy consensus in anonymization systems via game theory. In *Data and Applications Security and Privacy XXVI - 26th Annual IFIP WG 11.3 Conference (DBSec 2012)*, volume 7371 of *Lecture Notes in Computer Science*, pages 74–89. Springer, 2012.
3. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic databases. In *VLDB '02: Proceedings of the 28th international conference on Very Large Data Bases*, pages 143–154. VLDB Endowment, 2002.
4. H. E. Anderson. The privacy gambit: Toward a game theoretic approach to international data protection. *bepress Legal Series*, 2006.
5. P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter. Enterprise Privacy Authorization Language (EPAL 1.2). Technical report, IBM, 2003.
6. K. Barker, M. Askari, M. Banerjee, K. Ghazinour, B. Mackas, M. Majedi, S. Pun, and A. Williams. A data privacy taxonomy. In *BNCOD 26: Proceedings of the 26th British National Conference on Databases*, pages 42–54, Berlin, Heidelberg, 2009. Springer-Verlag.
7. R. Böhme, S. Koble, and T. U. Dresden. On the viability of privacy-enhancing technologies in a self-regulated business-to-consumer market: Will privacy remain a luxury good? In *WEIS 2007*, 2007.
8. G. Calzolari and A. Pavan. Optimal design of privacy policies. Technical report, Gremaq, University of Toulouse, 2001.
9. L. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. Reagle, M. Schunter, D. A. Stampley, and R. Wenning. The platform for privacy preferences 1.1 (p3p1.1) specification. World Wide Web Consortium, Note NOTE-P3P11-20061113, November 2006. W3C Recommendation.

10. L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. M. Reagle. The platform for privacy preferences 1.0 (p3p1.0) specification. World Wide Web Consortium, Recommendation REC-P3P-20020416, April 2002.
11. C. Dwork. Differential privacy. In *ICALP (2)*, pages 1–12, 2006.
12. N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *ICDE 2007*, pages 106–115, 2007.
13. T. Li and N. Li. On the tradeoff between privacy and utility in data publishing. In *KDD*, pages 517–526, New York, NY, USA, 2009. ACM.
14. G. Loukides and J. Shao. Data utility and privacy protection trade-off in k-anonymisation. In *PAIS 2008*, pages 36–45. ACM, 2008.
15. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1), Mar. 2007.
16. E. Mansfield and G. Yohe. *Microeconomics: Theory/Applications*. W.W. Norton, 2004.
17. G. R. Milne and M. E. Gordon. Direct mail privacy-efficiency trade-offs within an implied social contract framework. *Journal of Public Policy & Marketing*, 12(2):pp. 206–215, 1993.
18. T. Moses. eXtensible access control markup language TC v2.0 (XACML), 2005.
19. M. J. Osborne. *An Introduction to Game Theory*, chapter 8,9,16. Oxford University Press, USA, August 2003.
20. S. Preibusch. Implementing privacy negotiations in e-commerce. In *APWeb 2006*, volume 3841 of *Lecture Notes in Computer Science*, pages 604–615. Springer, 2006.
21. E. Singer, N. A. Mathiowetz, and M. P. Couper. The impact of privacy and confidentiality concerns on survey participation: The case of the 1990 u.s. census. *The Public Opinion Quarterly*, 57(4):pp. 465–482, 1993.
22. M. Sramka, R. Safavi-Naini, J. Denzinger, and M. Askari. A practice-oriented framework for measuring privacy and utility in data sanitization systems. In *Proceedings of the 2010 EDBT/ICDT Workshops*, EDBT '10, pages 27:1–27:10. ACM, 2010.
23. V. Staden and M. Olivier. Purpose organisation. In *Proceedings of the fifth annual Information Security South Africa (ISSA) Conference*, 2005.
24. L. Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.
25. W. van Staden and M. Olivier. Using purpose lattices to facilitate customisation of privacy agreements. In C. Lambrinoudakis, G. Pernul, and A. Tjoa, editors, *Trust, Privacy and Security in Digital Business*, volume 4657 of *Lecture Notes in Computer Science*, pages 201–209. Springer Berlin / Heidelberg, 2007.
26. M. P. Zielinski and M. S. Olivier. On the use of economic price theory to find the optimum levels of privacy and information utility in non-perturbative microdata anonymisation. *Data Knowl. Eng.*, 69(5):399–423, 2010.