HARVARD UNIVERSITY, COMPUTER SCIENCE DEPARTMENT DEPARTMENT

Efficiency in Two-Candidate Elections with Differential Privacy

Perry A. Green, K. Nathaniel Tucker, Matthew Warshauer

May 27, 2013

This paper will cover the topic of differential privacy in mechanism design. We will provide an intuitive understanding of privacy and why it is important in the larger realm of mechanism design: explaining how differential privacy fits with the other goals of mechanism design, truthfulness and efficiency. We will formalize these goals and examine how various mechanisms have previously tried to achieve them and what trade-offs they garnered in result. We run a variety of simulations to investigate the specific trade-off between privacy and efficiency in 2-candidate elections. Lastly, we try to find the privacy guarantees that optimize efficiency in circumstances where a mechanism designer can encourage more people to vote by increasing privacy.

1 INTRODUCTION

We will principally be viewing privacy through the lens of a 2-candidate election, however, these algorithms can be used for mechanisms in the realms of meeting scheduling, VCG (Vickrey Clarke Groves) auctions, digital goods auctions, etc.

In the realm of elections, what does privacy mean? For our purposes we assume that one participant or a third party "observer" will have the votes of each participant sans one, and will know the outcome. Thus, given the votes of everybody but one, and the outcome, can our "observer" deduce what your vote is? So our measure of privacy is: given the situation that can reveal the most, with what certainty can our "observer" ascertain your vote?

Consider the following example of a simple 2-candidate election with $2n+1$ voters. The situation that can reveal the most about a single voter is the case where n vote one way, and n vote the other. Thus we have one voter as the deciding vote. In this case our observer can exactly tell how that deciding voter swayed (just by looking at the outcome). Thus the privacy cost for a simple 2-candidate majority rules election is infinity, or the highest cost.

A situation where privacy is at its lowest cost, would be a mechanism where given all of the votes, the mechanism would assign the winner randomly. In this case, the "observer" would with no certainty be able to determine the choice of any voter. But here we can easily see that our mechanism would only give the desired result (majority wins) fifty percent of the time, we would say that this mechanism has low efficiency.

The goal must then be to balance these two conflicting goals, by sacrificing some probability of achieving the efficient outcome, in order to ensure modest but not complete privacy for the participating agents.

2 BACKGROUND

2.1 NOTATION

Throughout the paper we use the following notation for mechanisms and their properties:

- A number n of players. (eg. n voters in the election).
- A set Θ of player types. (eg. $\Theta = A, B$, where $\theta_i \in \Theta$ will indicate which of the two candidates is preferred by voter $i \in [n]$). We assume that an agent's action, if they participate in the election, is to just vote for their preferred candidate. So, we do not distinguish between types and actions for agents.
- A set of O outcomes. (eg. $O = A, B$, where the outcome indicates which candidate wins).
- An outcome function $M : \Theta^n \to O$ that determines an outcome given players' actions. (eg. using majority voting procedure to determine who wins).
- Player-specific utility functions $U_i : \Theta \times O \to \mathbb{R}$.

2.2 TRUTHFULNESS

Definition 1 (Truthfulness). $\forall i, \theta_i \in \Theta, \theta'_i \in \Theta, \theta_{-i}$

$$U_i(\theta_i, M(\theta_i, \theta_{-i})) \geq U_i(\theta_i, M(\theta'_i, \theta_{-i}))$$

A standard majority vote would be truthful, as in changing one's vote to a less preferred candidate will never grant our player more utility (in fact it either does not affect the outcome, or decreases our player's utility).

However there are looser definitions of truthfulness that we will encounter on our preview of differential privacy and mechanism design. For randomized mechanisms, defined as $M : \Theta^n \times R \to O$, where R is the probability space from which our mechanism draws its randomness. We will write: $M(\theta)$ to denote our sampling of r from R and evaluating $M(\theta; r)$. Thus a looser definition of truthfulness for randomized mechanisms is truthfulness in expectation:

Definition 2 (Truthfulness in Expectation). $\forall i, \theta_i \in \Theta, \theta'_i \in \Theta, \theta_{-i}$

$$E[U_i(\theta_i, M(\theta_i, \theta_{-i}))] \geq E[U_i(\theta_i, M(\theta'_i, \theta_{-i}))]$$

We also have a bit stronger notion called universal truthfulness:

Definition 3 (Universal Truthfulness). $\forall i, \theta_i \in \Theta, \theta'_i \in \Theta, \theta_{-i} r \in R$

$$U_i(\theta_i, M(\theta_i, \theta_{-i}; r)) \geq U_i(\theta_i, M(\theta'_i, \theta_{-i}; r))$$

This means that for any randomness that can be generated, $M(\theta; r)$ is a deterministic truthful mechanism.

Most likely one of the loosest relaxations is Approximate Truthfulness, where the mechanism will be within some bound of optimal:

Definition 4 (Approximate Truthfulness). $\forall i, \theta_i \in \Theta, \theta'_i \in \Theta, \theta_{-i}$

$$E[U_i(\theta_i, M(\theta_i, \theta_{-i}))] \leq e^{\epsilon} \times E[U_i(\theta_i, M(\theta'_i, \theta_{-i}))]$$

2.3 EFFICIENCY

In the context of 2-candidate elections, a result is efficient if it is consistent with the preferences of the majority of the voters. In contexts where some voters choose not to participate, their preferences are still considered when determining the welfare-maximizing result.

Here we will define our overall utility function as $\bar{U}(\Theta^n \times M)$, thus we can define a relaxation of efficiency requirement used in some papers with this notion:

Definition 5 ($\delta(n)$-efficient). $\forall \theta \in \Theta$

$$E_M[\bar{U}(\Theta^n, M)] \geq \max_{m \in \mathbb{M}} \bar{U}(\Theta^n, m) - \delta(n)$$

Where $\mathbb{M}$ is the set of all mechanisms. This will ensure that we are within some preset bound (based on number of individuals in the election) of the maximal utility.

2.4 Privacy

In general many of the papers will refer to differential privacy. We know that differentially private algorithms by definition are insensitive to individuals' inputs, meaning: a change in a single individual's input only has a small change in the outcome (we will see formalizations of this further on). Thus, one can easily see that one such of these mechanisms is innately "approximately truthful," or in other words reporting untruthfully can only provide a small gain (which it turns out we can bound) in a player's utility. Differential privacy is formalized below:

Definition 6 (ϵ-differentially Private)**.** $\forall \theta_{-i} \in \Theta^{n-1}, o \in O$

$$\max_{\theta',\theta'' \in \Theta} \frac{Pr[M(\theta'_i, \theta_{-i}) = o]}{Pr[M(\theta''_i, \theta_{-i}) = o]} \leq e^{\epsilon}$$

For ease of consideration later, we define the max ratio in the definition of differential privacy as the privacy ratio. For agent i it is the ratio of the most likely action he took given all other agent's actions and the observed outcome from the mechanism and the least likely action he might have taken given the same things. An infinite privacy ratio would mean a total lack of privacy. A privacy ratio of 1 would mean total privacy for the agent. A privacy guarantee can therefore be thought of as fixing the maximum possible privacy ratio under the mechanism. However, this must be converted to an ϵ value to fit into the notation, so if guaranteeing a max privacy ratio of x then a mechanism is $\epsilon = ln(x)$ differentially private. It is important to remember that higher privacy ratio guarantees actually mean weaker privacy for the agents.

One can also loosen this definition of privacy further to ϵ, η)-differentially private. This allowed later works to use exact truthfulness with private mechanisms:

Definition 7 ((ϵ, η)-differentially Private)**.** $\forall \theta_{-i} \in \Theta^{n-1}, o \in O$

$$\max_{\theta',\theta'' \in \Theta} \frac{Pr[M(\theta'_i, \theta_{-i}) = o]}{Pr[M(\theta''_i, \theta_{-i}) = o]} \leq e^{\epsilon} + \eta$$

Here we think of η as being $o(1)$.

3 Previous Works

We would overall like to pull together the corpus of works done on mechanism design that incorporates differential privacy. We will cover a variety of works and end with Chen et alius (1) as the mechanism design that we will use for subsequent experiments.

3.1 Overview

The first work that effectively brought together mechanism design and differential privacy was by McSherry and Talwar (2), where they used differential privacy as a tool for mechanism design. They study several unlimited supply auction problems, and provide a new mechanisms for digital goods auctions, attribute auctions, and auctions with arbitrary structural constraints on the prices, all using differential privacy in order to facilitate their results.

Specifically they looked at the case of an unlimited supply pricing problem, where an unlimited supply of goods must be made available to all at a single price, and they added laplacian noise according to a generalization of a differential privacy mechanism.

Their mechanism selects a price p such that the revenue is maximized and modifies that price with noise $\mathscr{E}(\Theta^n)$ which is represented with a uniform multiplier by a factor of the optimal price.

Generally they show that mechanisms with differential privacy are approximate dominant strategy under arbitrary player utility functions, are automatically resilient to coalitions, and easily allow repeatability. However, the purpose of the paper was not to incorporate differential privacy, but rather to solve the mechanism design problem of digital goods auction.

But again, as expressed above, approximate truthfulness, may not be the most satisfactory solution concept (as one can imagine, make the margins big enough and everything becomes approximately truthful). So while differential privacy can guarantee that a player will gain arbitrarily little by lying, it also makes the gain from telling the truth arbitrarily small. Nissim et alius even show that is some cases misreporting becomes a dominant strategy equilibrium (3). They conclude the paper by modifying some of the mechanisms of McSherry and Talwar (2) to provide truthfulness, but sacrifice differential privacy. Again the paper does not fundamentally seek to promote privacy, but rather seeks to solve a problem with mechanism design.

While it may seem up to this point that there exists a trade-off between differential privacy and truthfulness, Xiao remedied this deficiency to an extent and constructed mechanisms that simultaneously achieve truthfulness and differential privacy (4). However the model, in achieving exact truthfulness, loosened privacy concerns from ϵ-differentially private to (ϵ,η)-differentially private, and thus was able to achieve exact truthfulness. Xiao's model allows the player to only weakly prefer telling the truth and in certain scenarios lying does not reduce the player's utility at all. Also the paper fails to incorporate privacy into a player's utility. Regardless the work is foundational.

Finally, Chen et alius (1) incorporated privacy directly into the player's utility functions. The paper creates three mechanisms that are universally truthful, ϵ-differentially private, and optimal as n approaches infinity.

3.2 Summary

Paper	Truthfulness	Efficient	Private
McSherry and Talwar (2)	Approximate Truthfulness	$\delta(n)$-efficient	ϵ-Differentially Private
Nissim et alius (3)	Truthful	$\delta(n)$-efficient	Not Private
Xiao (4)	Truthful	$\delta(n)$-efficient	(ϵ,η)-Differentially Private
Chen et alius (1)	Universally Truthful	efficient in limit	ϵ-differentially private

4 Extensions and Analysis

A mechanism presented in Chen's paper for 2-Candidate elections is the basis for our extensions. It achieved the very attractive results of, universal truthfulness, efficient in the limit, and ϵ-differential privacy.

The mechanism is as follows:

1. Input a profile $\theta \in \{A, B\}^n$ and a privacy
2. Choose $r \in \mathbb{Z}$ from a discrete Laplace distribution, $Pr[r = k] \propto exp(-\epsilon|k|)$
3. Decide election by $\#\{i : \theta_i = A\} - \#\{i : \theta_i = B\} \geq r$, output A, otherwise B.

4.1 Tradeoffs Between Privacy Ratio Guarantee and Efficiency

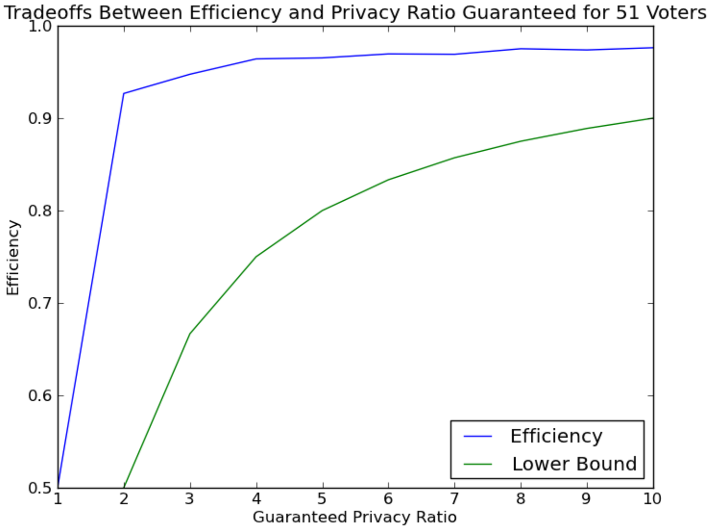
From Chen's paper, we have that, if i is the number of voters for the winning candidate under the mechanism, and j is the number of voters for the candidate with the most votes, that:

$$P(i \leq j + \Delta) < e^{-\epsilon\Delta}$$

By setting $\Delta = 1$, we find the probability of an inefficient outcome is bounded by $e^{-\epsilon}$. Following out convention of referring to efficiency for a given privacy ratio x, we conclude that the theoretical guarantee for the efficiency rate is:

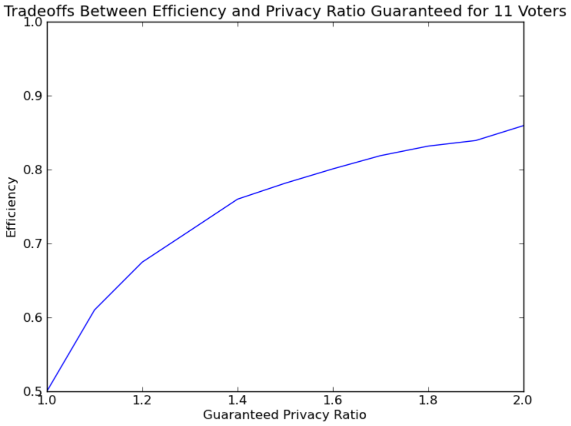
$$1 - e^{-\epsilon} = 1 - e^{-\log(x)} = 1 - \frac{1}{x}$$

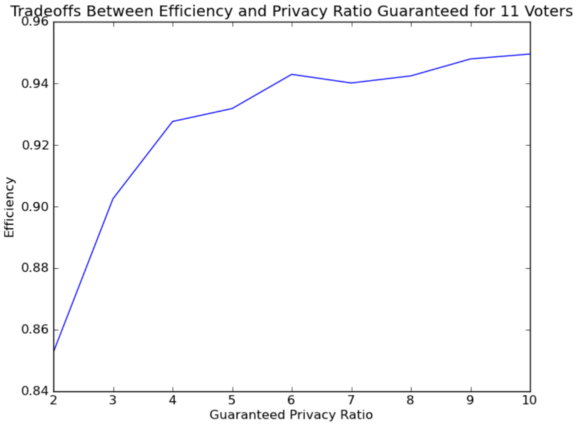
We confirmed this result in the following graph. For a 51-voter two-candidate election we see that the expected efficiency is well above the bound.

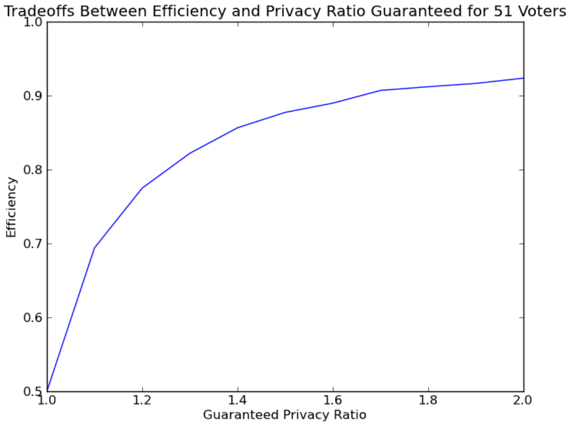


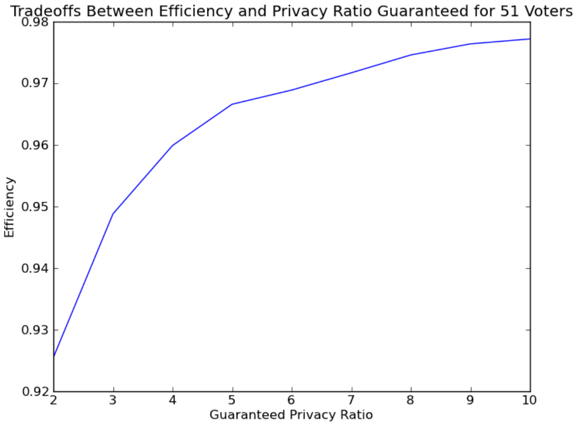
A quick note on simulations: For each vertex on each graph over nine thousand elections were simulated with the appropriate numbers of voters and voter's candidate preferences assigned as Bernoulli(.5).

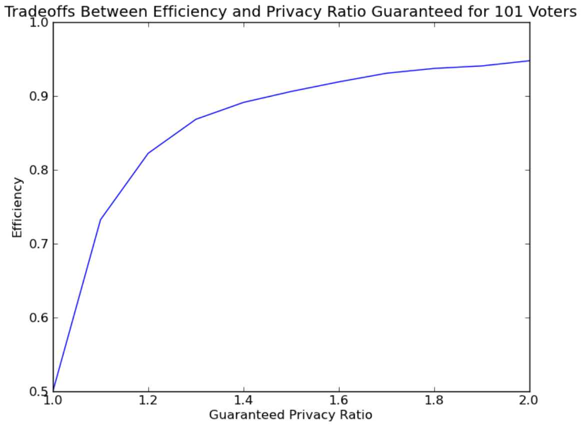
We simulated tens of thousands of 2-candidate elections with varying numbers of voters to explore the shape of the efficiency curve over variations in the privacy ratio guaranteed. As is hinted at from the bound graph above, the scale of the behavior is very different over $(1,2]$ than over $[2,10]$. Thus, we present those domains separately and with different scales. Note that a privacy ratio of 1 cannot actually be achieved so the behavior shown on the graphs is for arbitrarily close to 1 instead. We do this analysis for 11, 51, and 101 voters to illustrate how increasing the number of voters increases efficiency for all guaranteed privacy ratios. This will be explored more later.

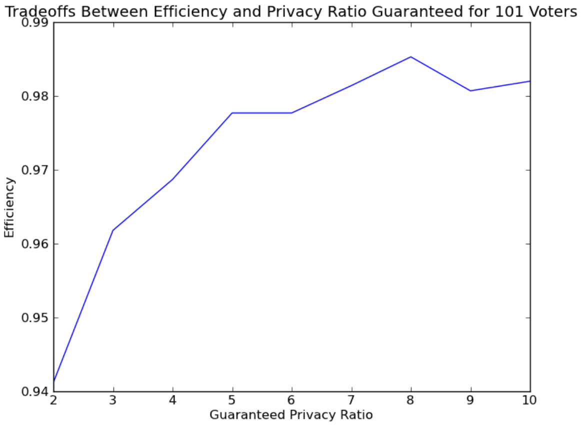






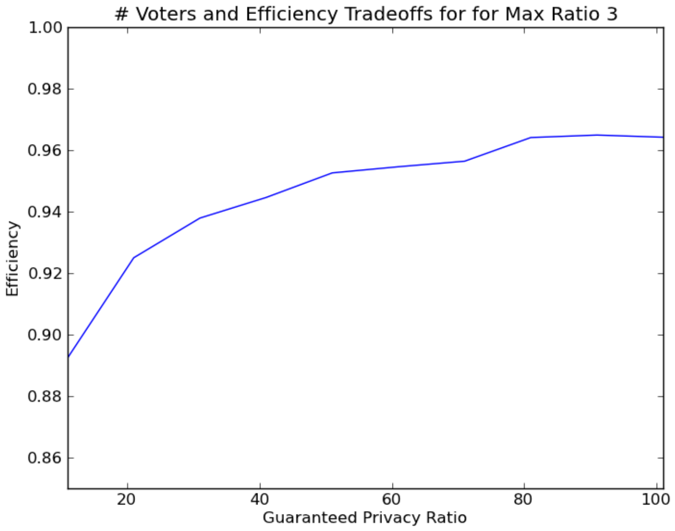






From these we learn that, even though there is interesting behavior over both domains, the vast majority of the efficiency comes from the (1,2] domain of ratios. Efficiency appears to exhibit diminishing marginal returns to increases in the privacy ratio guaranteed. Even though this mechanism continues to offer strong theoretical results, they become weak in practice because the ever higher ratios needed to achieve small gains in efficiency offer only very weak guarantees of privacy to agents. The importance of this of course depends on agents' preferences over their privacy - a topic to which we will return later.

We explore more directly the effect on efficiency of varying the number of voters. The following graph is the efficiency over tens of thousands of simulations of two-candidate elections with the privacy ratio guaranteed fixed to 3 and the number of voters ranging from 11 to 101 by 10.



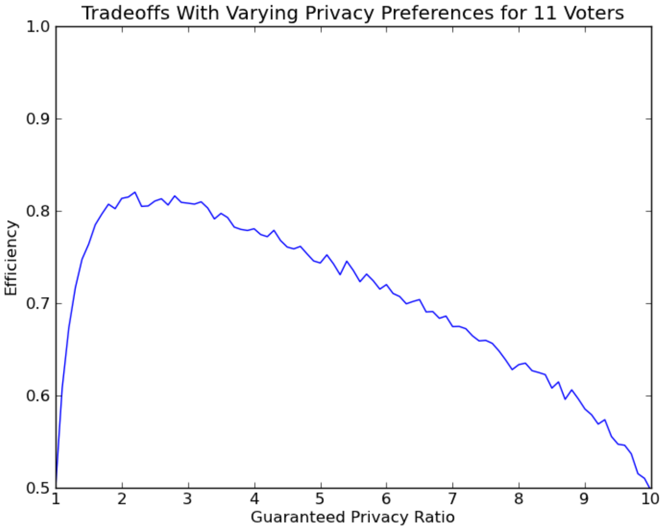
As expected, the efficiency increases in the number of voters. This is consistent with what we saw above. More voters means that the added noise is less likely to change the outcome of the election because the margin in the real votes is more likely to be greater than the magnitude of the noise.

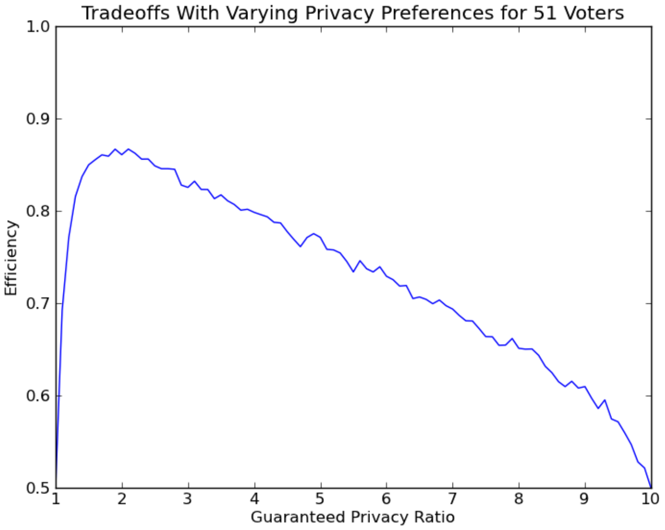
4.2 Mechanisms Seeking Efficient Outcomes with Varyingly Private Voters

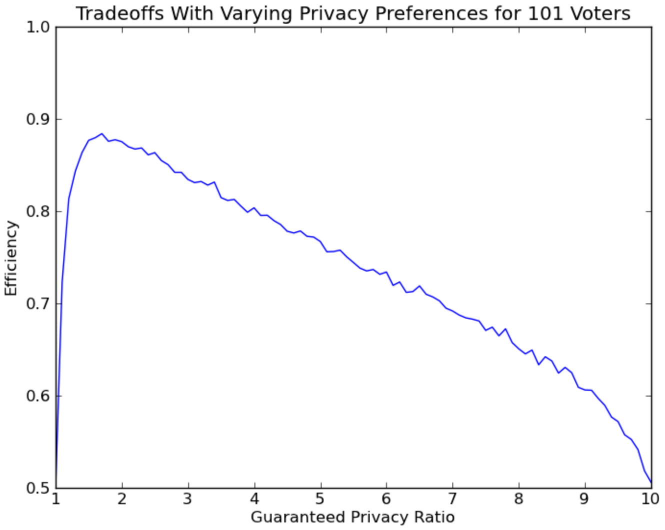
Now Chen's paper used the introduction of an agent's utility of privacy to consider the question of what a mechanism designer must do in order to get agents to vote. However, we propose that the goal of the mechanism designer is not directly to maximize the number of voters, but rather to maximize the efficiency of the election mechanism. In other words, the designer prioritizes that her mechanism output the true majority preference of the voters and that this is not the same as trying to maximize the number of agents who vote. We know increasing the number of voters increases efficiency. And decreasing the strength of the privacy guarantee by increasing the max privacy ratio increases efficiency (by adding less noise) but at the cost of driving out voters who are not comfortable with the privacy level. Given these two competing forces there is some tradeoff between weakening privacy to get less noise along with fewer voters and strengthening privacy to get more noise along with more voters.

More formally, assume all agent's have some maximum privacy ratio that must be guaranteed to them in order to ensure their participation in the election. We assume that they derive negative infinite utility from participating in an election which has an insufficient privacy guarantee. If the guarantee is sufficient than there is no utility effect from privacy. Further assume that they derive some positive utility from their preferred candidate winning the election so that, conditional upon the election being sufficiently private, they always vote truthfully.

We might imagine that the the sufficient privacy ratio for a particular agent is drawn from a uniform distribution over the range $(1,10)$. Now we graph the simulated election results for 11, 51, and 101 voters varying guaranteed privacy ratio.


Tradeoffs With Varying Privacy Preferences for 11 Voters
1.0
0.9
0.8
0.7
0.6
0.5
Efficiency
1
2
3
4
5
6
7
8
9
10
Guaranteed Privacy Ratio


Tradeoffs With Varying Privacy Preferences for 51 Voters
1.0
0.9
0.8
0.7
0.6
0.5
Efficiency
1
2
3
4
5
6
7
8
9
10
Guaranteed Privacy Ratio



This is first interesting as a proof of concept that there is in fact an ideal peak in the tradeoff between increasing efficiency by decreasing the amount of noise and by increasing participation by agents in the election. This peak is at the guaranteed privacy ratio that the mechanism designer should use in order to maximize expected efficiency under these circumstances.

Note that at a privacy ratio close to 1 nearly everyone participates but there is large noise so the outcome is essentially random. At a privacy ratio close to 10 very few participate and, though there is little noise, the small number of voters means that the outcome is again essentially random because the sample is unlikely to be representative. This explains why both ends of each graph are at .5. Second, we observe that the quick growth and peak near 2 aligns with the previous observation that most change in expected efficiency comes over the range $(1,2]$.

Now we consider how these structurally similar graphs vary with different numbers of potential voters. As expected, increasing the number of voters raises efficiency at all points along the graph. There are two ways in which we might imagine that increasing the number of potential voters influences the location of the peak.

1. More potential voters means a smaller percentage of the voters need to turn out in order to get a representative sample and so the designer can profitably offer weaker privacy ratios (shift to the right) in order to minimize noise without losing sample representativeness.

2. More potential voters means the cost of the noise (which is independent of number of potential voters) becomes increasingly unimportant when compared to the benefit from getting more voters to turnout (shift to the left)

Based the the clear observed shifts to the left in the graphs we support the second statement.

5 Further Work

The perspective taken in the end of our paper, as a mechanism designer seeking to maximize efficiency in the face of agents with privacy demands, is as far as we can tell a new application of differential privacy to mechanism design. As such, there is a lot of room for further work along these lines. For example, Chen's paper includes a general VCG-like mechanism with differential privacy. With further time, we would have liked to try and adapt our experiments to mechanisms like these which are not two-candidate elections. In addition, there are further steps that can be taken to investigate the two-candidate election problem more deeply. For example, we only tried sampling the privacy demands of voters from a uniform distribution. It might be interesting to see how the results change when sampled from a normal distribution, or to try to come up with theoretical results for arbitrary distributions.

6 Bibliography

1. Chen, Yiling, et al. "Truthful mechanisms for agents that value privacy." arXiv preprint arXiv:1111.5472 (2011).

2. Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In FOCS, pages 94âĂŞ 103. IEEE Computer Society, 2007.

3. Kobbi Nissim, Rann Smorodinsky, and Moshe Tennenholtz. Approximately optimal mechanism design via differential privacy. CoRR, abs/1004.2888, 2010.

4. David Xiao. Is privacy compatible with truthfulness? Technical Report 2011/005, Cryptology ePrint Archive, 2011.

5. Yael Onn, et. al., Privacy in the Digital Environment , Haifa Center of Law & Technology, (2005) pp. 1-12