Approximate Privacy: Foundations and Quantification

JOAN FEIGENBAUM, Yale University
AARON D. JAGGARD, Rutgers University
and MICHAEL SCHAPIRA, Yale University and UC Berkeley

The proliferation of online sensitive data about individuals and organizations makes concern about the *privacy* of these data a top priority. There have been many formulations of privacy and, unfortunately, many negative results about the feasibility of maintaining privacy of sensitive data in realistic networked environments. We formulate communication-complexity-based definitions, both worst-case and average-case, of a problem's *privacy-approximation ratio*. We use our definitions to investigate the extent to which approximate privacy is achievable in a number of standard problems: the 2^{nd}-price Vickrey auction, Yao's millionaires problem, the public-good problem, and the set-theoretic disjointness and intersection problems.

For both the 2^{nd}-price Vickrey auction and the millionaires problem, we show that not only is perfect privacy impossible or infeasibly costly to achieve, but even *close approximations* of perfect privacy suffer from the same lower bounds. By contrast, if the inputs are drawn uniformly at random from $\{0, \ldots, 2^k - 1\}$, then, for both problems, simple and natural communication protocols have privacy-approximation ratios that are linear in k (*i.e.*, logarithmic in the size of the input space). We also demonstrate tradeoffs between privacy and communication in a family of auction protocols.

We show that the privacy-approximation ratio provided by any protocol for the disjointness and intersection problems is necessarily exponential (in k). We also use these ratios to argue that one protocol for each of these problems is significantly fairer than the others we consider (in the sense of relative effects on the privacy of the different players).

Categories and Subject Descriptors: K.4.1 [**Computers and Society**]: Public Policy Issues—*privacy*

General Terms: Economics, Security

Additional Key Words and Phrases: Approximate privacy, bisection auction

ACM Reference Format:
Joan Feigenbaum, Aaron D. Jaggard, and Michael Schapira, YYYY. Approximate Privacy: Foundations and Quantification. *ACM Trans. Algor.* V, N, Article A (YYYY), 38 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

This work was presented in preliminary form at the 2010 ACM Conference on Electronic Commerce [Feigenbaum et al. 2010a]. Feigenbaum was supported in part by NSF grants 0331548, 0534052, and 0716223 and IARPA grant FA8750-07-0031. Jaggard was supported in part by NSF grants 0751674, 0753492, and 1018557. Schapira was supported by NSF grant 0331548.
Author's addresses: J. Feigenbaum, Department of Computer Science, Yale University, P.O. Box 208285, 51 Prospect Street, New Haven, CT 06520-8285; A. D. Jaggard, DIMACS Center, Rutgers University, (Current address) Formal Methods Section (Code 5543), U.S. Naval Research Laboratory, 4555 Overlook Ave. SW, Washington, DC 20375 (this work does not necessarily reflect the views of the U.S. Government); M. Schapira, Departments of Computer Science, Yale University and University of California, Berkeley, (Current address) The Selim and Rachel Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Ross Building, The Edmond J. Safra Campus, 91904 Jerusalem, Israel.

DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Increasing use of computers and networks in business, government, recreation, and almost all aspects of daily life has led to a proliferation of online sensitive data about individuals and organizations. Consequently, the study of privacy has become a top priority in many disciplines. Computer scientists have contributed many formulations of the notion of *privacy-preserving computation* that have opened new avenues of investigation and shed new light on some well studied problems.

One good example of a new avenue of investigation opened by concern about privacy can be found in auction design, which was our original motivation for this work. Traditional auction theory is a central research area in Economics, and one of its main questions is how to incent bidders to behave truthfully, *i.e.*, to reveal private information that auctioneers need in order to compute optimal outcomes. More recently, attention has turned to the complementary goal of enabling bidders *not* to reveal private information that auctioneers do *not* need in order to compute optimal outcomes. The importance of bidders' privacy, like that of algorithmic efficiency, has become clear now that many auctions are conducted online, and Computer Science has become at least as relevant as Economics.

Our approach to privacy is based on communication complexity. Although originally motivated by agents' privacy in mechanism design, our definitions and tools can be applied to distributed function computation in general. Because perfect privacy can be impossible or infeasibly costly to achieve, we investigate approximate privacy. Specifically, we formulate both worst-case and average-case versions of the *privacy-approximation ratio* of a function f in order to quantify the amount of privacy that can be maintained by parties who supply sensitive inputs to a distributed computation of f. We also study the tradeoff between privacy preservation and communication complexity.

Our points of departure are the work of Chor and Kushilevitz [Chor and Kushilevitz 1991] on characterization of privately computable functions and that of Kushilevitz [Kushilevitz 1992] on the communication complexity of private computation. Starting from the same place, Bar-Yehuda *et al.* [Bar-Yehuda et al. 2006] also provided a framework in which to quantify the amount of privacy that can be maintained in the computation of a function and the communication cost of achieving it. Their definitions and results are significantly different from the ones we present here (see discussion in Sec. 1.4); as explained in Section 8 below, a precise characterization of the relationship between their formulation and ours is an interesting direction for future work.

1.1. Our Approach

Consider an auction of a Bluetooth headset with 2 bidders, 1 and 2, in which the auctioneer accepts bids ranging from \$0 to \$7 in \$1 increments. Each bidder i has a private value $x_i \in \{0, \ldots, 7\}$ that is the maximum he is willing to pay for the headset. The item is sold in a 2^{nd}-price Vickrey auction, *i.e.*, the *higher* bidder gets the item (with ties broken in favor of bidder 1), and the price he pays is the *lower* bid. The demand for privacy arises naturally in such scenarios [Naor et al. 1999]: In a straightforward protocol, the auctioneer receives sealed bids from both bidders and computes the outcome based on this information. Say, *e.g.*, that bidder 1 bids \$3, and bidder 2 bids \$6. The auctioneer sells the headset to bidder 2 for \$3. It would not be at all surprising however if, in subsequent auctions of headsets in which bidder 2 participates, the same auctioneer set a reservation price of \$5. This could be avoided if the auction protocol allowed the auctioneer to learn the fact that bidder 2 was the highest bidder (something he needs to know in order to determine the outcome) but did not entail the full revelation of 2's private value for the headset.

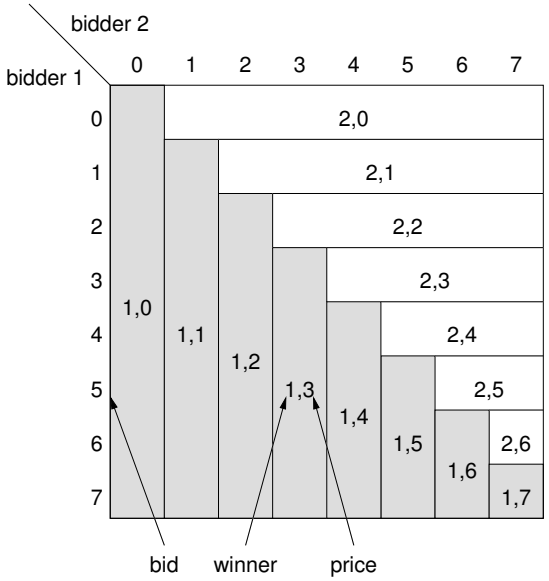

Fig. 1. The minimal knowledge requirements for 2^{nd}-price auctions

Observe that, in some cases, revelation of the exact private information of the highest bidder is necessary. For example, if $x_1 = 6$, then bidder 2 will win only if $x_2 = 7$. In other cases, the revelation of a lot of information is necessary, *e.g.*, if bidder 1's bid is 5, and bidder 2 outbids him, then x_2 must be either 6 or 7. An auction protocol is said to achieve *perfect objective privacy* if the auctioneer learns nothing about the private information of the bidders that is not needed in order to compute the result of the auction. Figure 1 illustrates the information the auctioneer *must* learn in order to determine the outcome of the 2^{nd}-price auction described above. Observe that the auctioneer's failure to distinguish between two potential pairs of inputs that belong to different rectangles in Fig. 1 implies his inability to determine the winner or the price the winner must pay. Also observe, however, that the auctioneer need not be able to distinguish between two pairs of inputs that belong to the same rectangle.

Using the "minimal knowledge requirements" described in Fig. 1, we can now characterize a perfectly (objective) privacy-preserving auction protocol as one that induces this *exact* partition of the space of possible inputs into subspaces in which the inputs are indistinguishable to the auctioneer. Unfortunately, perfect privacy is often hard or even impossible to achieve. For 2^{nd}-price auctions, Brandt and Sandholm [Brandt and Sandholm 2008] show that *every* perfectly private auction protocol has exponential communication complexity. This provides the motivation for our definition of *privacy-approximation ratio*: We are interested in whether there is an auction protocol that achieves "good" privacy guarantees without paying such a high price in computational efficiency. We no longer insist that the auction protocol induce a partition of inputs *exactly* as in Fig. 1 but rather that it "approximate" the optimal partition well. We define two kinds of privacy-approximation ratio (PAR): *worst-case* PAR and *average-case* PAR.

The worst-case PAR of a protocol P for the 2^{nd}-price auction is defined as the maximum ratio between the size of a set S of indistinguishable inputs in Fig. 1 and the size of a set of indistinguishable inputs induced by P that is contained in S. If a protocol is perfectly privacy preserving, these sets are always the same size, and so the worst-

case PAR is 1. If, however, a protocol fails to achieve perfect privacy, then at least one "ideal" set of indistinguishable inputs *strictly* contains a set of indistinguishable inputs induced by the protocol. In such cases, the worst-case PAR will be strictly higher than 1.

Consider, *e.g.*, the sealed-bid auction protocol in which both bidders reveal their private information to the auctioneer, who then computes the outcome. Obviously, this naive protocol enables the auctioneer to distinguish between every two pairs of private inputs, and so each set of indistinguishable inputs induced by the protocol contains exactly one element. The worst-case PAR of this protocol is therefore $\frac{8}{1} = 8$. (If bidder 2's value is 0, then in Fig. 1 the auctioneer is unable to determine which value in $\{0, \ldots, 7\}$ is x_1. In the sealed bid auction protocol, however, the auctioneer learns the exact value of x_1.) The *average-case* PAR is a natural Bayesian variant of this definition: We now assume that the auctioneer has knowledge of some market statistics, in the form of a probability distribution over the possible private information of the bidders. PAR in this case is defined as the average ratio and not as the maximum ratio as before.

Thus, intuitively, PAR captures the effect of a protocol on the privacy (in the sense of indistinguishability from other inputs) afforded to protocol participants—it indicates the factor by which, in the worst case or on average, using the protocol to compute the function, instead of just being told the output, reduces the number of inputs from which a given input cannot be distinguished. To formalize and generalize the above intuitive definitions of PAR, we make use of machinery from communication-complexity theory. Specifically, we use the concepts of *monochromaticity* and *tilings* to make formal the notions of sets of indistinguishable inputs and of the approximability of privacy. We discuss other notions of approximate privacy in Section 8.

1.2. Our Findings

We present both upper and lower bounds on the privacy-approximation ratio for both the millionaires problem and 2^{nd}-price auctions with 2 bidders. Our analysis of these two environments takes place within Yao's 2-party communication model [Yao 1979], in which the private information of each party is a k-bit string, representing a value in $\{0, \ldots, 2^k - 1\}$. In the millionaires problem, the two parties (the millionaires) wish to keep their private information hidden *from each other*. We refer to this goal as the preservation of *subjective* privacy. In electronic-commerce environments, each party (bidder) often communicates with the auctioneer via a secure channel, and so the aim in the 2^{nd}-price auction is to prevent *a third party* (the auctioneer), who is unfamiliar with any of the parties' private inputs, from learning "too much" about the bidders. This goal is referred to, in this paper, as the preservation of *objective* privacy.

Informally, for both the 2^{nd}-price Vickrey auction and the millionaires problem, we obtain the following results: We show that not only is perfect privacy impossible or infeasibly costly to achieve, but even close approximations of perfect privacy suffer from the same lower bounds. By contrast, we show that, if the values of the parties are drawn uniformly at random from $\{0, \ldots, 2^k - 1\}$, then, for both problems, simple and natural communication protocols have privacy-approximation ratios that are linear in k (*i.e.*, logarithmic in the size of the space of possible inputs). We conjecture that this improved PAR is achievable for *any* probability distribution. The correctness of this conjecture would imply that, no matter what beliefs the protocol designer may have about the parties' private values, a protocol that achieves reasonable privacy guarantees exists.

Importantly, our results for the 2^{nd}-price Vickrey auction are obtained by proving a more general result for a large family of protocols for single-item auctions, termed *bounded-bisection auctions*, that contains both the celebrated ascending-price English auction and the class of bisection auctions [Fujishima et al. 1999; Grigorieva et al.

2006; Grigorieva et al. 2007; Herings et al. 2009]. We are thus able to quantify a tradeoff between communication complexity and the effect on privacy within this family of protocols.

We show that our results for the millionaires problem also extend to the classic economic problem of *provisioning a public good*, by observing that, in terms of privacy-approximation ratios, the two problems are, in fact, equivalent. We also consider a truthful variant of the public-good problem and show that this can have a bounded PAR (depending on how the cost of the good grows).

We then apply our PAR framework to two classic set-theoretic problems: the intersection problem (in which party 1's input is a set S_1, party 2's input is a set S_2, and the goal of the protocol is to compute $S_1 \cap S_2$) and its decision version disjointness (in which $f(S_1, S_2) = 1$ if $S_1 \cap S_2 = \emptyset$, and $f(S_1, S_2) = 0$ otherwise). From both the privacy perspective and the communication-complexity perspective, these are extremely natural problems to study. The intersection problem has served as a motivating example in the study of privacy-preserving computation for decades; in a typical application, two organizations wish to compute the set of members that they have in common without disclosing to each other the people who are members of only one of the organizations. The disjointness problem plays a central role in the theory and application of communication complexity, where the fact that $n + 1$ bits of communication are required to test disjointness of two subsets of $\{1, \ldots, n\}$ is used to prove many worst-case lower bounds.

In applying our PAR framework to the disjointness and intersection problems, we consider three natural protocols that apply to both problems. We compute the objective and subjective PARs for all three protocols for both problems. The objective and subjective PARs are exponential in all cases, but we show that the protocol that is intuitively the best is quantifiably (and significantly) more fair than the others in the sense described below; to do this, we consider the ratios of the subjective PARs (as described in Sec. 3.3) and argue that this captures some intuitive sense of fairness.

1.3. Related Work and Follow-Up Work: Defining Privacy-Preserving Computation

1.3.1. Communication-Complexity-Based Privacy Formulations. As explained above, the privacy work of Bar-Yehuda *et al.* [Bar-Yehuda et al. 2006] and the work presented in this paper have common ancestors in [Chor and Kushilevitz 1991; Kushilevitz 1992]. Similarly, the work of Brandt and Sandholm [Brandt and Sandholm 2008] uses Kushilevitz's formulation to prove an exponential lower bound on the communication complexity of privacy-preserving 2^{nd}-price Vickrey auctions. We elaborate on the relation of our work to that of Bar-Yehuda *et al*. [Bar-Yehuda et al. 2006] in Sec. 1.4.

Similarly to [Bar-Yehuda et al. 2006; Chor and Kushilevitz 1991; Kushilevitz 1992], our work focuses on the two-party deterministic communication model. We view our results as first step in a more general research agenda, outlined in Sec. 8.

There are many formulations of privacy-preserving computation, both exact and approximate, that are not based on the definitions and tools in [Chor and Kushilevitz 1991; Kushilevitz 1992]. We now briefly review some of them and explain how they differ from ours.

1.3.2. Secure, Multiparty Function Evaluation. The most extensively developed approach to privacy in distributed computation is that of *secure, multiparty function evaluation* (SMFE). Indeed, to achieve agent privacy in algorithmic mechanism design, which was our original motivation, one could, in principle, simply start with a strategyproof mechanism and then have the agents themselves compute the outcome and payments using an SMFE protocol. However, as observed by Brandt and Sandholm [Brandt and

Sandholm 2008], these protocols fall into two main categories, and both have inherent disadvantages from the point of view of mechanism design:

— *Information-theoretically* private protocols, the study of which was initiated by Ben-Or, Goldwasser, and Wigderson [Ben-Or et al. 1988] and Chaum, Crépeau, and Damgård [Chaum et al. 1988], rely on the assumption that a constant fraction of the agents are "honest" (or "obedient" in the terminology of distributed algorithmic mechanism design [Feigenbaum and Shenker 2002]), *i.e.*, that they follow the protocol perfectly even if they know that doing so will lead to an outcome that is not as desirable to them as one that would result from their deviating from the protocol; clearly, this assumption is antithetical to the main premise of mechanism design, which is that all agents will behave strategically, deviating from protocols when and only when doing so will improve the outcome from their points of view;
— Multiparty protocols that use *cryptography* to achieve privacy, the study of which was initiated by Yao [Yao 1982; 1986], rely on (plausible but currently unprovable) complexity-theoretic assumptions. Often, they are also very communication-intensive (see, *e.g.*, [Brandt and Sandholm 2008] for an explanation of why some of the deficiencies of the Vickrey auction cannot be solved via cryptography). Moreover, sometimes the deployment cryptographic machinery is infeasible (over the years, many cryptographic variants of the current interdomain routing protocol, BGP, were proposed, but not deployed due to the infeasibility of deploying a global Internet-wide PKI infrastructure and the real-time computational cost of verifying signatures). For some mechanisms of interest, efficient cryptographic protocols have been obtained (see, *e.g.*, [Dodis et al. 2000; Naor et al. 1999]).

In certain scenarios, the demand for *perfect* privacy preservation cannot be relaxed. In such cases, if the function cannot be computed in a privacy-preserving manner without the use of cryptography, there is no choice but to resort to a cryptographic protocol. There is an extensive body of work on cryptography-based identity protocols, and we are not offering our notion of PAR as an extension of that work. (In fact, the framework described here might be applied to SMFE protocols by replacing indistinguishability by computational indistinguishability. However, this does not appear to yield any new insights.)

However, in other cases, we argue that privacy preservation should be regarded as one of *several design goals*, alongside low computational/communication complexity, protocol simplicity, incentive-compatibility, and more. (See, *e.g.*, [Muthukrishnan 2009].) Therefore, it is necessary to be able to *quantify* privacy preservation in order to understand the tradeoffs among the different design goals, and obtain "reasonable" (but not necessarily perfect) privacy guarantees. Our PAR approach continues the long line of research about information-theoretic notions of privacy, initiated by Ben-Or *et al*. and by Chaum *et al*. Regardless of the above argument, we believe that information-theoretic formulations of privacy and approximate privacy are also natural to consider in their own right.

1.3.3. Private Approximations and Approximate Privacy. In this paper, we consider protocols that compute exact results but preserve privacy only approximately. One can also ask what it means for a protocol to compute approximate results in a privacy-preserving manner; indeed, this question has also been studied [Beimel et al. 2008; Feigenbaum et al. 2006; Halevi et al. 2001], but it is unrelated to the questions we ask here. Similarly, definitions and techniques from *differential privacy* [Dwork 2006] (see also [Ghosh et al. 2009]), in which the goal is to add noise to the result of a database query in such a way as to preserve the privacy of the individual database records (and hence protect the data subjects) but still have the result convey nontrivial information,

are inapplicable to the problems that we study here. Ghosh and Roth [Ghosh and Roth 2011] have used the differential-privacy framework to consider the problem of selling privacy, which leads to the use of an auction for privacy; in addition to using different techniques, that problem is different from that of how much privacy is provided by the auction protocol itself (which is one of the problems to which we apply our framework). Ghosh and Roth show that no mechanism can compensate agents sufficiently to guarantee incentive compatibility if the values that agents place on their privacy are unbounded; moreover, they do not attempt to hide the privacy valuations from the mechanism. McGregor *et al*. [McGregor et al. 2010] study the setting of a database distributed between two parties who run a protocol such that each party's view is a differentially private function of the other's data. They proved lower bounds on the accuracy of such protocols, in part using connections between differential privacy and communication complexity. Nissim *et al*. [Nissim et al. 2012] design *privacy-aware* mechanisms that are incentive-compatible with respect to all agents whose privacy valuations are below a certain threshold and differentially private with respect to all other agents.

1.4. Relation to the Work of Bar-Yehuda *et al.*

While there are certainly some parallels between the work here and the Bar-Yehuda *et al*. work [Bar-Yehuda et al. 2006], there are significant differences in what the two frameworks capture. Specifically:

(1) The results in [Bar-Yehuda et al. 2006] deal with what can be learned by a party who knows one of the inputs. By contrast, our notion of objective PAR captures the effect of a protocol on privacy with respect to an external observer who does not know any of the players private values.

(2) More importantly, the framework of [Bar-Yehuda et al. 2006] does not address the size of monochromatic regions. As illustrated by the following example, the ability to do so is necessary to capture the effects of protocols on interesting aspects of privacy that are captured by our definitions of PAR.

Consider the function $f : \{0, \ldots, 2^n - 1\} \times \{0, \ldots, 2^n - 1\} \to \{0, \ldots, 2^{n-2}\}$ defined by $f(x, y) = floor(\frac{x}{2})$ if $x < 2^{n-1}$ and $f(x, y) = 2^{n-2}$ otherwise. Consider the following two protocols for f: in P, player 1 announces his value x if $x < 2^{n-1}$ and otherwise sends 2^{n-1} (which indicates that $f(x, y) = 2^{n-2}$); in Q, player 1 announces $floor(\frac{x}{2})$ if $x < 2^n - 1$ and x if $x = 2^n - 1$. Observe that each protocol induces $2^{n-1} + 1$ rectangles.

Intuitively, the effect on privacy of these two protocols is different. For half of the inputs, P reduces by a factor of 2 the number of inputs from which they are indistinguishable while not affecting the indistinguishability of the other inputs. Q does not affect the indistinguishability of the inputs affected by P, but it does reduce the number of inputs indistinguishable from a given input with $x \geq 2^{n-1}$ by at least a factor of 2^{n-2}.

Our notion of PAR is able to capture the different effects on privacy of the protocols P and Q. (The average-case objective PARs are constant and exponential in n, respectively.) By contrast, the three quantifications of privacy from [Bar-Yehuda et al. 2006]—$\mathcal{I}_c$, $\mathcal{I}_i$, and $\mathcal{I}_{c-i}$—do not distinguish between these two protocols; we now sketch the arguments for this claim.

For each protocol, any function h for which the protocol is weakly h-private must take at least $2^{n-1}+1$ different values. This bound is tight for both P and Q. Thus, $\mathcal{I}_c$ cannot distinguish between the effects of P and Q on f.

The number of rectangles induced by P that intersect each row and column equals the number induced by Q. Considering the geometric interpretation of I_P and I_Q in Lemma 7 of [Bar-Yehuda et al. 2006], as well as the discussion in Sec. VII.A of [Bar-

Yehuda et al. 2006], we see that $\mathcal{I}_i$ and $\mathcal{I}_{c-i}$ (when restricted to a single protocol, which erases the distinction between them) cannot distinguish between the effects of P and Q on f.

In Sec. 8 we give examples that show the objective (*i.e.*, from the perspective of an outside observer) analogues of the information-theoretic measures used in defining $\mathcal{I}_i$ and $\mathcal{I}_{c-i}$ can disagree with PAR on whether the effect of two protocols is similar and even on which protocol is more private.

1.5. Follow-Up Work

Subsequent to our preliminary work, several papers adopted our approximate-privacy framework to further explore the research directions outlined in this work [Ada et al. 2012], to analyze other objective functions and communication protocols of interest [Comi et al. 2012], and to study efficiency and privacy tradeoffs in mechanism design [Sui and Boutilier 2011].

1.6. Paper Outline

In the next section, we review and expand upon the connection between perfect privacy and communication complexity. We present our formulations of approximate privacy, both worst case and average case, in Section 3; we present our PAR results on second-price auctions, the millionaires problem, and public-good problems in Sections 5 and 4. Section 6 gives formal definitions of the disjointness and intersection problems, describes the protocols that we consider for these problems, and gives a summary and discussion of our related PAR results. Section 7 gives the full statements and proofs of our PAR results for disjointness and intersection, respectively. In Sec. 8, we discuss other approaches that might be taken to the problem of quantifying partial privacy; we also discuss various directions for future research in this area.

2. PERFECT PRIVACY AND COMMUNICATION COMPLEXITY

We now briefly review Yao's model of two-party communication and notions of objective and subjective perfect privacy; see Kushilevitz and Nisan [Kushilevitz and Nisan 1997] for a comprehensive overview of communication complexity theory. Note that we only deal with *deterministic* communication protocols. Our definitions can be extended to randomized protocols.

2.1. Two-Party Communication Model

There are two parties, 1 and 2, each holding a k-bit *input string*. The input of party i, $x_i \in \{0,1\}^k$, is the *private information* of i. The parties communicate with each other in order to compute the value of a function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}^t$. The two parties alternately send messages to each other. In communication round j, one of the parties sends a bit q_j that is a function of that party's input and the history $(q_1, \ldots, q_{j-1})$ of previously sent messages. We say that a bit is *meaningful* if it is not a constant function of this input and history and if, for every meaningful bit transmitted previously, there is some combination of input and history for which the bit differs from the earlier meaningful bit. Non-meaningful bits (*e.g.*, those sent as part of protocol-message headers) are irrelevant to our work here and will be ignored. A *communication protocol* dictates, for each party, when it is that party's turn to transmit a message and what message he should transmit, based on the history of messages and his value.

A communication protocol P is said to compute f if, for every pair of inputs (x_1, x_2), it holds that $P(x_1, x_2) = f(x_1, x_2)$. As in [Kushilevitz 1992], the last message sent in a protocol P is assumed to contain the value $f(x_1, x_2)$ and therefore may require up to t bits. The *communication complexity* of a protocol P is the maximum, over all input pairs, of the number of bits transmitted during the execution of P.

Any function $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}^t$ can be visualized as a $2^k \times 2^k$ matrix with entries in $\{0,1\}^t$, in which the rows represent the possible inputs of party 1, the columns represent the possible inputs of party 2, and each entry contains the value of f associated with its row and column inputs. This matrix is denoted by $A(f)$.

Definition 2.1 (*Regions, partitions*). A *region* in a matrix A is any subset of entries in A (not necessarily a submatrix of A). A *partition* of A is a collection of disjoint regions in A whose union equals A.

Definition 2.2 (*Monochromaticity*). A region R in a matrix A is called *monochromatic* if all entries in R contain the same value. A *monochromatic partition* of A is a partition all of whose regions are monochromatic.

Of special interest in communication complexity are specific kinds of regions and partitions called rectangles, and tilings, respectively:

Definition 2.3 (*Rectangles, Tilings*). A *rectangle* in a matrix A is a submatrix of A. A *tiling* of a matrix A is a partition of A into rectangles.

Definition 2.4 (*Refinements*). A partition or tiling $P_1(f)$ of a matrix $A(f)$ is said to be a *refinement* of another partition or tiling $P_2(f)$ of $A(f)$ if every region in $P_1(f)$ is contained in some region in $P_2(f)$.

Monochromatic rectangles and tilings are an important concept in communication-complexity theory, because they are linked to the execution of communication protocols. Every communication protocol P for a function f can be thought of as follows:

(1) Let R and C be the sets of row and column indices of $A(f)$, respectively. For $R' \subseteq R$ and $C' \subseteq C$, we will abuse notation and write $R' \times C'$ to denote the submatrix of $A(f)$ obtained by deleting the rows not in R' and the columns not in C'.

(2) While $R \times C$ is not monochromatic:

— One party $i \in \{1,2\}$ sends a single bit q (whose value is based on x_i and the history of communication).

— If $i = 1$, q indicates whether 1's value is in one of two disjoint sets R_1, R_2 whose union equals R. If $x_1 \in R_1$, both parties set $R = R_1$. If $x_1 \in R_2$, both parties set $R = R_2$.

— If $i = 2$, q indicates whether 2's value is in one of two disjoint sets C_1, C_2 whose union equals C. If $x_2 \in C_1$, both parties set $C = C_1$. If $x_2 \in C_2$, both parties set $C = C_2$.

(3) One of the parties sends a last message (consisting of up to t bits) containing the value in all entries of the monochromatic rectangle $R \times C$.

Observe that, for every pair of private inputs (x_1, x_2), P terminates at some monochromatic rectangle in $A(f)$ that contains (x_1, x_2). We refer to this rectangle as *the monochromatic rectangle induced by P for* (x_1, x_2). We refer to the tiling that consists of all rectangles induced by P (for all pairs of inputs) as *the monochromatic tiling induced by P*.

Remark 2.5. There are monochromatic tilings that cannot be induced by communication protocols. For example, observe that the tiling in Fig. 2 (which is essentially an example from [Kushilevitz 1992]) has this property. The first input-dependent bit that is sent in the protocol will divide at least one monochromatic region into two parts. If this bit is sent by party 1, whose possible input values correspond to the rows, then at least one of the bottom-left and top-right regions shown will be divided unnecessarily. (It is possible that the function takes the same value on these regions, so that they in fact form a single monochromatic region; that single region would still be divided.)

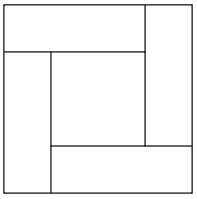

Fig. 2. A tiling that cannot be induced by any communication protocol [Kushilevitz 1992]

2.2. Perfect Privacy

Informally, we say that a two-party protocol is *perfectly privacy-preserving* if the two parties (or a third party observing the communication between them) cannot learn more from the execution of the protocol than the value of the function the protocol computes. (These definition can be extended naturally to protocols involving more than two participants.)

Formally, let P be a communication protocol for a function f. The *communication string* passed in P is the concatenation of all the messages $(q_1, q_2 \ldots)$ sent in the course of the execution of P. Let $s_{(x_1,x_2)}$ denote the communication string passed in P if the inputs of the parties are (x_1, x_2). We are now ready to define perfect privacy. The following two definitions handle privacy from the point of view of a party i that does not want the other party (who is, of course, familiar not only with the communication string, but also with *his own* value) to learn more than necessary about i's private information. We say that a protocol is perfectly private with respect to party 1 if 1 never learns more about party 2's private information than necessary to compute the outcome.

Definition 2.6 (*Perfect privacy with respect to 1*). [Chor and Kushilevitz 1991; Kushilevitz 1992] P is *perfectly private with respect to party* 1 if, for every x_1, x_2, x_2' such that $f(x_1, x_2) = f(x_1, x_2')$, it holds that $s_{(x_1,x_2)} = s_{(x_1,x_2')}$.

Informally, Def. 2.6 says that party 1's knowledge of the communication string passed in the protocol and his knowledge of x_1 do not aid him in distinguishing between two possible inputs of 2. Similarly:

Definition 2.7 (*Perfect privacy with respect to 2*). [Chor and Kushilevitz 1991; Kushilevitz 1992] P is *perfectly private with respect to party* 2 if, for every x_1, x_1', x_2 such that $f(x_1, x_2) = f(x_1', x_2)$, it holds that $s_{(x_1,x_2)} = s_{(x_1',x_2)}$.

Observation 2.8. For any function f, the protocol in which party i reveals x_i and the other party computes the outcome of the function is perfectly private with respect to i.

Definition 2.9 (*Perfect subjective privacy*). P achieves *perfect subjective privacy* if it is perfectly private with respect to both parties.

The following definition considers a different form of privacy—privacy from *a third party* that observes the communication string but has no *a priori* knowledge about the private information of the two communicating parties. We refer to this notion as *objective privacy*.

Definition 2.10 (*Perfect objective privacy*). P achieves *perfect objective privacy* if, for every two pairs of inputs (x_1, x_2) and (x_1', x_2') such that $f(x_1, x_2) = f(x_1', x_2')$, it holds that $s_{(x_1,x_2)} = s_{(x_1',x_2')}$.

	0	1
0	1	0
1	0	1

Fig. 3. Matrix for $f(x_1, x_2) = 1 \oplus x_1 \oplus x_2$ (with x_1 shown in the left column and x_2 in the top row) illustrating the differences between subjective and objective privacy.

Figure 3 illustrates the difference between perfect subjective and objective privacy. The function is $f(x_1, x_2) = 1 \oplus x_1 \oplus x_2$, and we assume that different input pairs produce different communication strings. For every value of x_1, the conditions of Def. 2.6 are trivially satisfied; in terms of the matrix, for each row, the inputs in the row that produce the same function value also have the same communication string. This also holds for fixed values of x_2 and the conditions of Def. 2.7 or, equivalently, for fixed columns of the matrix. By contrast, because $(0, 0)$ and $(1, 1)$ produce the same output value, perfect *objective* privacy would require that those inputs as well as $(0, 1)$ and $(1, 0)$ produce the same communication strings (including output values).[1]

Kushilevitz [Kushilevitz 1992] was the first to point out the interesting connections between perfect privacy and communication-complexity theory. Intuitively, we can think of any monochromatic rectangle R in the tiling induced by a protocol P as a set of inputs that are *indistinguishable* to a third party. This is because, by definition of R, for any two pairs of inputs in R, the communication string passed in P must be the same. Hence we can think of the privacy of the protocol in terms of the tiling induced by that protocol.

Ideally, every two pairs of inputs that are assigned the same outcome by a function f will belong to the same monochromatic rectangle in the tiling induced by a protocol for f. This observation enables a simple characterization of perfect privacy-preserving mechanisms.

Definition 2.11 (*Ideal monochromatic partitions*). A monochromatic region in a matrix A is said to be a *maximal monochromatic region* if no monochromatic region in A properly contains it. *The ideal monochromatic partition* of A is made up of the maximal monochromatic regions.

Observation 2.12. For every possible value in a matrix A, the maximal monochromatic region that corresponds to this value is unique. This implies the uniqueness of the ideal monochromatic partition for A.

Observation 2.13 (*A characterization of perfectly privacy-preserving protocols*). A communication protocol P for f is perfectly privacy-preserving iff the monochromatic tiling induced by P is the ideal monochromatic partition of $A(f)$. This holds for all of the above notions of privacy.

3. PRIVACY-APPROXIMATION RATIOS

Unfortunately, perfect privacy should not be taken for granted. As shown by our results, in many environments, perfect privacy can be either impossible or very costly (in terms of communication complexity) to obtain. To measure a protocol's effect on privacy, relative to the ideal—but perhaps impossible to implement—computation of the outcome of a problem, we introduce the notion of *privacy-approximation ratios* (PARs).

[1]This check will be useful in subsequent examples. We may look at the matrix for the function as in Fig. 3; if there are two input pairs (i, j) and (i', j') that produce the same function value, the function cannot be computed perfectly privately if the inputs (i, j') and (i', j) do not also produce this same function value. The tiling from Fig. 2 shows that satisfying this is necessary but not sufficient, however.

3.1. Worst-Case PARs

For any communication protocol P for a function f, we denote by $R^P(x_1, x_2)$ the monochromatic rectangle induced by P for (x_1, x_2). We denote by $R^I(x_1, x_2)$ the monochromatic region containing $A(f)_{(x_1,x_2)}$ in the ideal monochromatic partition of $A(f)$. Intuitively, $R^P(x_1, x_2)$ is the set of inputs that are indistinguishable from (x_1, x_2) to P. $R^I(x_1, x_2)$ is the set of inputs that *would be* indistinguishable from (x_1, x_2) if perfect privacy were preserved. We wish to asses how far one is from the other. The size of a region R, denoted by $|R|$, is the cardinality of R, *i.e.*, the number of inputs in R.

We can now define worst-case *objective* PAR as follows:

Definition 3.1 (*Worst-case objective PAR of* P). The *worst-case objective privacy-approximation ratio* of communication protocol P for function f is

$$\alpha = \max_{(x_1,x_2)} \frac{|R^I(x_1, x_2)|}{|R^P(x_1, x_2)|}.$$

We say that P is α-*objective-privacy-preserving in the worst case*.

Definition 3.2 (i-*partitions*). The 1-*partition* of a region R in a matrix A is the set of disjoint rectangles $R_{x_1} = \{x_1\} \times \{x_2 \; s.t. \; (x_1, x_2) \in R\}$ (over all possible inputs x_1). 2-*partitions* are defined analogously.

Intuitively, given any region R in the matrix $A(f)$, if party i's actual private information is x_i, then i can use this knowledge to eliminate all the parts of R other than R_{x_i}. Hence, the other party should be concerned not with R but rather with the i-partition of R.

Definition 3.3 (i-*induced tilings*). The i-*induced tiling* of a protocol P is the refinement of the tiling induced by P obtained by i-partitioning each rectangle in it.

Definition 3.4 (i-*ideal monochromatic partitions*). The i-*ideal monochromatic partition* is the refinement of the ideal monochromatic partition obtained by i-partitioning each region in it.

For any communication protocol P for a function f, we use $R_i^P(x_1, x_2)$ to denote the monochromatic rectangle containing $A(f)_{(x_1,x_2)}$ in the i-induced tiling for P. We denote by $R_i^I(x_1, x_2)$ the monochromatic rectangle containing $A(f)_{(x_1,x_2)}$ in the i-ideal monochromatic partition of $A(f)$.

Definition 3.5 (*Worst-case PAR of* P *with respect to* i). The *worst-case privacy-approximation ratio with respect to* i of communication protocol P for function f is

$$\alpha = \max_{(x_1,x_2)} \frac{|R_i^I(x_1, x_2)|}{|R_i^P(x_1, x_2)|}.$$

We say that P is α-*privacy-preserving with respect to* i *in the worst case*.

Definition 3.6 (*Worst-case subjective PAR of* P). The *worst-case subjective privacy-approximation ratio* of communication protocol P for function f is the maximum of the worst-case privacy-approximation ratio with respect to each party.

Definition 3.7 (*Worst-case PAR*). The *worst-case objective (subjective) PAR for a function* f is the minimum, over all protocols P for f, of the worst-case objective (subjective) PAR of P.

3.2. Average-Case PARs

As we shall see below, it is also useful to define an average-case version of PAR. As the name suggests, the average-case objective PAR is the *average* ratio between the size of the monochromatic rectangle containing the private inputs and the corresponding region in the ideal monochromatic partition.

Definition 3.8 (*Average-case objective PAR of* P). Let D be a probability distribution over the space of inputs. The *average-case objective privacy-approximation ratio* of communication protocol P for function f is

$$\alpha = E_D \left[\frac{|R^I(x_1, x_2)|}{|R^P(x_1, x_2)|} \right].$$

We say that P is α*-objective privacy-preserving in the average case with distribution* D (or *with respect to* D).

We define average-case PAR with respect to i analogously, and average-case subjective PAR as the maximum over all players i of the average-case PAR with respect to i. We define the *average-case objective (subjective) PAR for a function* f as the minimum, over all protocols P for f, of the average-case objective (subjective) PAR of P.

In computing the average-case PAR (either objective or subjective) with respect to the uniform distribution, we may simplify the previous expressions for PAR values. If each player's value space has k bits, then the average-case objective PAR with respect to the uniform distribution equals

$$\mathsf{PAR}(k) = \sum_{(x_1,x_2)} \frac{1}{2^{2k}} \frac{|R^I(x_1, x_2)|}{|R^P(x_1, x_2)|},$$

where the sum is over all pairs (x_1, x_2) in the value space. We may combine all of the terms corresponding to points in the same protocol-induced rectangle to obtain

$$\mathsf{PAR}(k) = \sum_S \frac{|S|}{2^{2k}} \frac{|R^I(S)|}{|S|} = \frac{1}{2^{2k}} \sum_S |R^I(S)|, \tag{1}$$

where the sums are now over protocol-induced rectangles S. Note also that the average-case PAR with respect to i and with respect to the uniform distribution is obtained by replacing $R^I(S)$ with $R^I_i(S)$ in Eq. 1.

It may seem that the occurrences of set cardinality in the quantity considered in Def. 3.8 should be replaced by the probability measure of the regions in question. However, as we discuss in Sec. 8.1, such a definition is unable to distinguish between examples that should be viewed as having very different levels of privacy; by contrast, the definition that we consider here is able to distinguish between such cases.

We note that extending a function and protocol to a larger domain can significantly change the value of the protocol's PAR. If the domain is enlarged sufficiently and the expanded protocol is sufficiently private (or input-revealing), the extended protocol's PAR will be dominated by the contribution from the extended domain.

3.3. Ratios of Subjective PARs

We may also be interested in another quantity related to PAR. Given some protocol P for a function f, let $\mathsf{PAR}^i_D(k)$ be the average-case subjective PAR of P with respect to protocol participant i and distribution D on the k-bit input space. We then let

$$\mathsf{PAR}^{\mathsf{max}}_D(k) = \max_i \mathsf{PAR}^i_D(k) \qquad \text{and} \qquad \mathsf{PAR}^{\mathsf{min}}_D(k) = \min_i \mathsf{PAR}^i_D(k),$$

where the $\max$ and $\min$ are taken over all protocol participants. We then define the *ratio of (average-case with respect to D) subjective PARs* to be

$$\frac{\mathsf{PAR}_D^{\mathsf{max}}(k)}{\mathsf{PAR}_D^{\mathsf{min}}(k)} \geq 1.$$

Intuitively, in a two-participant protocol, this captures how much greater a negative effect the protocol P can have on one participant than on the other participant. The average-case subjective PAR of a protocol P identifies the maximum effect that P can have on the privacy with respect to a participant. However, it does not capture whether this effect is similar for both players, and in fact this effect can be quite different. Below we show that, for both the disjointness and intersection problems, there are protocols that have exponentially large subjective PARs; for some protocols, the subjective PAR with respect to one player is exponentially larger than that with respect to the other player, while for one protocol for each problem, the subjective PARs with respect to the different players differ only by a constant (asymptotic) factor. We argue that this is an important distinction and that the ratio of average-case subjective PARs captures some intuitive notion of the fairness of the protocol. If a protocol has a much larger PAR with respect to player 2 than with respect to player 1, an agent might agree to participate in a protocol run only if he is assigned the role of player 2 (so that he learns much more about the other player than the other player learns about him). Thus, from the perspective of the protocol implementer who needs to induce participation, protocols with small ratios of average-case subjective PARs would likely be more desirable.

4. 2^{ND}-PRICE AUCTIONS: BOUNDS ON PARS

In this section, we present upper and lower bounds on the privacy-approximation ratios for the 2^{nd}-price Vickrey auction and discuss the tradeoff between average-case PAR and communication complexity for a family of auction protocols. We start with a formal statement of the problem.

4.1. Problem Specification and Summary of Results

2^{nd}*-price Vickrey auction.*. A single item is offered to 2 bidders, each with a private value for the item. The auctioneer's goal is to allocate the item to the bidder with the highest value. The fundamental technique in mechanism design for inducing truthful behavior in single-item auctions is Vickrey's 2^{nd}-price auction [Vickrey 1961]: Allocate the item to the highest bidder, and charge him the second-highest bid.

Definition 4.1 (2ND-PRICE AUCTION$_k$).
Input: $x_1, x_2 \in \{0, \ldots, 2^k - 1\}$ (each represented by a k-bit string)
Output: the identity of the party with the higher value, *i.e.*, $\arg\max_{i \in \{1,2\}} x_i$ (breaking ties lexicographically), and the private information of the other party.

Brandt and Sandholm [Brandt and Sandholm 2008] show that a perfectly privacy-preserving communication protocol exists for 2ND-PRICE AUCTION$_k$. Specifically, perfect privacy is obtained via the *ascending-price English auction*: Start with a price of $p = 0$ for the item. In each time step, increase p by 1 until one of the bidders indicates that his value for the item is less than p (in each step first asking bidder 1 and then, if necessary, asking bidder 2). At that point, allocate the item to the other bidder for a price of $p-1$. If p reaches a value of $2^k - 1$ (that is, the values of both bidders are $2^k - 1$) allocate the item to bidder 1 for a price of $2^k - 1$.

Moreover, it is shown in [Brandt and Sandholm 2008] that the English auction is essentially *the only* perfectly privacy-preserving protocol for 2ND-PRICE AUCTION$_k$. Thus, perfect privacy requires, in the worst-case, the transmission of $\Omega(2^k)$ bits. $2k$

Table I. Average-case PARs (with respect to the uniform distribution) for 2ND-PRICE AUCTION$_k$

	Avg.-Case Obj. PAR	Avg.-Case Subj. PAR
English Auction	1	1
BISECTION AUCTION$_{g(k)}$	$\frac{g(k)+3}{2} - \frac{2^{g(k)}}{2^{k+1}} + \frac{1}{2^{k+1}} - \frac{1}{2^{g(k)+1}}$	$\frac{g(k)+5}{4} - \frac{1}{2^{g(k)+2}} + \frac{g(k)}{2^{k+2}}$
BISECTION AUCTION	$\frac{k}{2}+1$	$\frac{k+5}{4} + \frac{k-1}{2^{k+2}}$
Sealed-Bid Auction	$\frac{2^{k+1}}{3} + \frac{1}{3\cdot 2^k}$	$\frac{2^k}{3} + 1 - \frac{1}{3\cdot 2^k}$

bits suffice to compute the function without privacy because bidders can simply reveal their inputs. Can we obtain "good" privacy without paying such a high price in communication?

Table I summarizes the average-case PAR results (with respect to the uniform distribution) for 2ND-PRICE AUCTION$_k$ that we obtain in the rest of this section. As discussed in Sec. 4.4, BISECTION AUCTION$_{g(k)}$, which is parameterized by a function $g : \mathbb{Z}^+ \to \mathbb{Z}^{\geq 0}$, interpolates between the ascending-price English Auction ($g(k) = 0$) and the BISECTION AUCTION ($g(k) = k$), which is discussed in Sec. 4.2.

4.2. Objective-Privacy PARs

We now consider *objective privacy* for 2ND-PRICE AUCTION$_k$ (*i.e.*, privacy with respect to the auctioneer). The BISECTION AUCTION [Fujishima et al. 1999; Grigorieva et al. 2006; Grigorieva et al. 2007; Herings et al. 2009] for 2ND-PRICE AUCTION$_k$ is defined as follows: Start by asking each player whether his value lies in $[0, 2^{k-1})$ or in $[2^{k-1}, 2^k)$; continue this binary search until the players' answers differ, at which point we know which bidder has the higher value. (If the values do not differ, we will also discover this; in this case, award the item to bidder 1, who must pay the common value.) Then use another binary search on the interval that contains the value of the lower bidder in order to find his value. Truthfulness is a weakly dominant strategy in BISECTION AUCTION [Grigorieva et al. 2006; Grigorieva et al. 2007; Herings et al. 2009]. Furthermore, for every k and every other protocol for 2ND-PRICE AUCTION$_k$, the number of input pairs for which the BISECTION AUCTION requires at most k steps is at least as large as the number of pairs for which the other protocol requires at most k steps [Grigorieva et al. 2010, Thm. 4.7].

More generally, we refer to an auction protocol as a c-bisection auction, for a constant $c \in (0, 1)$, if in each step the interval R is partitioned into two disjoint subintervals: a lower subinterval of size $c|R|$ and an upper subinterval of size $(1 - c)|R|$. Hence, the BISECTION AUCTION is a c-bisection auction with $c = \frac{1}{2}$. We prove that no c-bisection auction for 2ND-PRICE AUCTION$_k$ obtains a subexponential objective PAR:

THEOREM 4.2 (A WORST-CASE LOWER BOUND FOR c-BISECTION AUCTIONS).
For any constant $c > \frac{1}{2^k}$, the c-bisection auction for 2ND-PRICE AUCTION$_k$ *has a worst-case PAR of at least $2^{\frac{k}{2}}$.*

PROOF. Consider the ideal monochromatic partition of 2ND-PRICE AUCTION$_k$ depicted for $k = 3$ in Fig. 1. Observe that, for perfect privacy to be preserved, it must be that bidder 2 transmits the first (meaningful) bit, and that this bit partitions the space of inputs into the leftmost shaded rectangle (the set $\{0, \ldots, 2^k - 1\} \times \{0\}$) and the rest of the value space (ignoring the rectangles depicted that further refine $\{0, \ldots, 2^k - 1\} \times \{1, \ldots, 2^k - 1\}$). What if the first bit is transmitted by player 2 and does not partition the space into rectangles in that way? We observe that any other partition of the space into two rectangles is such that, in the worst case, the privacy-approximation ratio is at least $2^{\frac{k}{2}}$ (for any value of c): If $c \leq 1 - 2^{-\frac{k}{2}}$, then the case in

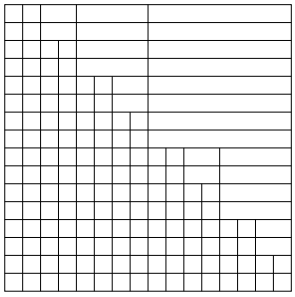

Fig. 4. The BISECTION-AUCTION-induced refinement of the 1-partition for 2ND-PRICE AUCTION$_k$ ($k = 4$)

which $x_1 = c2^k - 1$ gives us the lower bound. If, on the other hand, $c > 1 - 2^{-\frac{k}{2}}$, then the case that $x_1 = 0$ gives us the lower bound. Observe that such a bad PAR is also the result of bidder 1's transmitting the first (meaningful) bit. □

By contrast, as for THE MILLIONAIRES PROBLEM$_k$, reasonable privacy guarantees are achievable in the average case; this follows from general results below (Thm. 4.9 with $g(k) = k$).

THEOREM 4.3. **(The average-case objective PAR of the bisection auction)** The average-case objective PAR of the BISECTION AUCTION is $\frac{k}{2} + 1$ with respect to the uniform distribution.

We conjecture that, for any distribution, similarly good objective PAR can be achieved for the 2^{nd}-price Vickrey auction. Ada *et al*. have proved the analogue of this conjecture for the variant of PAR that they use [Ada et al. 2012].

CONJECTURE 4.4. *For any probability distribution D over the k-bit input space, the average-case objective PAR of* 2ND-PRICE AUCTION$_k$ *is linear in k.*

We note that the worst-possible approximation of objective privacy comes when each value in the space is in a distinct tile; this is the tiling induced by the sealed-bid auction. The resulting average-case privacy-approximation ratio is exponential in k.

Observation 4.5 (*Largest possible objective PAR*). The largest possible (for any protocol) average-case objective PAR with respect to the uniform distribution for 2ND-PRICE AUCTION$_k$ is $\frac{2}{3}2^k + \frac{1}{3}2^{-k}$.

4.3. Subjective Privacy PARs

We now look briefly at subjective privacy for 2ND-PRICE AUCTION$_k$. For subjective privacy with respect to 1, we start with the 1-partition for 2ND-PRICE AUCTION$_k$; Fig. 4 shows the refinement of the 1-partition induced by the BISECTION AUCTION for $k = 4$. Separately considering the refinement of the 2-partition for 2ND-PRICE AUCTION$_k$ by the BISECTION AUCTION, we have the following results.

Using more general results below, we may obtain the average-case PARs with respect to 1 and 2 (taking $g(k) = k$ in Thms. 4.13 and 4.14, respectively) for the bisection auction with respect to the uniform distribution. For 1, this is $\frac{k+3}{4} - \frac{k-1}{2^{k+2}}$; for 2, this is $\frac{k+5}{4} + \frac{k-1}{2^{k+2}}$, and we immediately have the following corollary:

COROLLARY 4.6. **(The average-case subjective PAR of the bisection auction)** The average-case subjective PAR of the BISECTION AUCTION with respect to the uni-

form distribution is

$$\frac{k+5}{4} + \frac{k-1}{2^{k+2}}.$$

As with objective privacy, we conjecture that, for any distribution, similarly good subjective PAR can be achieved for the 2^{nd}-price Vickrey auction.

CONJECTURE 4.7. *For any probability distribution D over the k-bit input space, the average-case subjective PAR of* 2ND-PRICE AUCTION$_k$ *is linear in k.*

Also as with objective privacy, the sealed-bid auction gives the largest possible average-case subjective PAR.

Observation 4.8 (*Largest possible subjective PAR*). The largest possible (for any protocol) average-case subjective PAR with respect to the uniform distribution for 2ND-PRICE AUCTION$_k$ is $\frac{2^k}{3} + 1 - \frac{1}{3 \cdot 2^k}$.

4.4. Bounded-Bisection Auctions

We now present a middle ground between the perfectly-private yet highly inefficient (in terms of communication) ascending English auction and the communication-efficient BISECTION AUCTION whose average-case objective PAR is linear in k (and is thus unbounded as k goes to infinity): We bound the number of bisections, using an ascending English auction to determine the outcome if it is not resolved by the limited number of bisections.

We define the BISECTION AUCTION$_{g(k)}$ as follows: Given an instance of 2ND-PRICE AUCTION$_k$, and an integer-valued function $g(k)$ such that $0 \leq g(k) \leq k$, run the BISECTION AUCTION as above but do at most $g(k)$ bisection operations. (Note that we will never do more than k bisections.) If the outcome is undetermined after $g(k)$ bisection operations, so that both players' values lie in an interval I of size $2^{k-g(k)}$, apply the ascending-price English auction to this interval to determine the identity of the winning bidder and the value of the losing bidder.

As $g(k)$ ranges from 0 to k, the BISECTION AUCTION$_{g(k)}$ ranges from the ascending-price English auction to the BISECTION AUCTION. If we allow a fixed, positive number of bisections ($g(k) = c > 0$), computations show that for $c = 1, 2, 3$ we obtain examples of protocols that do not provide perfect privacy but that do have bounded average-case objective PARs with respect to the uniform distribution. We wish to see if this holds for all positive c, determine the average-case objective PAR for general $g(k)$, and connect the amount of communication needed with the approximation of privacy in this family of protocols. The following theorem allows us to do these things.

THEOREM 4.9. *For the* BISECTION AUCTION$_{g(k)}$*, the average-case objective PAR with respect to the uniform distribution equals*

$$\frac{g(k)+3}{2} - \frac{2^{g(k)}}{2^{k+1}} + \frac{1}{2^{k+1}} - \frac{1}{2^{g(k)+1}}.$$

PROOF. Fix k, the number of bits used for bidding, and let $c = g(k)$ be the number of bisections; we have $0 \leq c \leq k$, and we let $i = k - c$. Figure 5 illustrates this tiling for $k = 4$, $c = 2$, and $i = 2$; note that the upper-left and lower-right quadrants have identical structure and that the lower-left and upper-right quadrants have no structure other than that of the ideal partition and the quadrant boundaries (which are induced by the first bisection operation performed).

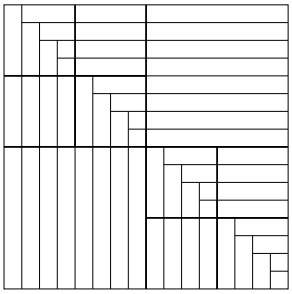

Fig. 5. Illustration for the proof of Thm. 4.9

Recall from Eq. 1 that

$$\mathsf{PAR}(k) = \frac{1}{2^{2k}} \sum_{S} |R(S)|, \tag{2}$$

gives the average-case objective PAR of a protocol with respect to the uniform distribution, where the sum is over protocol-induced rectangles S, and $R(S)$ denotes the ideal region containing S. Each ideal region in which bidder 1 wins is a rectangle of width 1 and height at most 2^k; each ideal region in which bidder 2 wins is a rectangle of height 1 and width strictly less than 2^k. For a protocol-induced rectangle S, let $j_S = 2^k - |R(S)|$. Let $a_{c,i}$ be the total number of tiles that appear in the tiling of the k-bit value space induced by the BISECTION AUCTION$_{g(k)}$ with $g(k) = c$, and let $b_{c,i} = \sum_S j_S$ (with this sum being over the protocol-induced tiles in this same partition). Then we may rewrite Eq. 2 as

$$\mathsf{PAR}_{c,i} = \frac{1}{2^{2k}} \sum_{S} (2^k - j_S) = \frac{a_{c,i} 2^k - b_{c,i}}{2^{2k}}. \tag{3}$$

(Note that Eq. 2 holds for general protocols; we now add the subscripts "c, i" to indicate the particular protocol whose PAR we are computing.) We now determine $a_{c,i}$ and $b_{c,i}$.

Considering the tiling induced by $c+1$ bisections of a $(c+i+1)$-bit space (which has $a_{c+1,i}$ total tiles), the upper-left and lower-right quadrants each contain $a_{c,i}$ tiles while the lower-left and upper-right quadrants (as depicted in Fig. 5) each contribute 2^{c+i} tiles, so $a_{c+1,i} = 2a_{c,i} + 2^{c+i+1}$. When there are no bisections, the i-bit value space has $a_{0,i} = 2^{i+1} - 1$ tiles, from which we obtain $a_{c,i} = 2^c \left(2^i(c+2) - 1\right)$. The sum of j_S over protocol-induced rectangles S in the upper-left quadrant is $b_{c,i}$. For a rectangle S in the lower-right quadrant, j_S equals 2^{c+i} plus $j_{S'}$, where S' is the corresponding rectangle in the upper-left quadrant; there are $a_{c,i}$ such S, so the sum of j_S over protocol-induced rectangles S in the upper-left quadrant is $b_{c,i} + a_{c,i} 2^{c+i}$. Finally, the sum of j_S over S in the lower-left quadrant equals $\sum_{h=0}^{2^{c+i}-1} h$ and the sum over S in the top-right quadrant equals $\sum_{h=1}^{2^{c+i}} h$. Thus, $b_{c+1,i} = 2b_{c,i} + a_{c,i} 2^{c+i} + 2^{2(c+i)}$; with $b_{0,i} = \sum_{h=0}^{2^i-1} h + \sum_{h=1}^{2^i-1} h$, we obtain $b_{c,i} = 2^{c+i-1} \left((1+2^c)\left(-1+2^i\right) + 2^{c+i} c\right)$. Rewriting Eq. 3, we obtain

$$\mathsf{PAR}_{c,i}(k) = \frac{c+3}{2} - \frac{2^c}{2^{c+i+1}} + \frac{1}{2^{c+i+1}} - \frac{1}{2^{c+1}}.$$

Recalling that $k = c + i$, the proof is complete. □

For the protocols corresponding to values of $g(k)$ ranging from 0 to k (ranging from the ascending-price English auction to the BISECTION AUCTION), we may thus relate

the amount of communication saved (relative to the English auction) to the effect of this on the PAR.

COROLLARY 4.10. *Let g be a function that maps non-negative integers to non-negative integers. Then the average-case objective PAR with respect to the uniform distribution for the* BISECTION AUCTION$_{g(k)}$ *is bounded if g is bounded and is unbounded if g is unbounded. We then have that the* BISECTION AUCTION$_{g(k)}$ *may require the exchange of $\Theta(k + 2^{k-g(k)})$ bits, and it has an average-case objective PAR of $\Theta(1 + g(k))$.*

Remark 4.11. Some of the sequences that appear in the proof above also appear in other settings. For example, the sequences $\{a_{0,i}\}_i$, $\{a_{1,i}\}_i$, and $\{a_{2,i}\}_i$ are slightly shifted versions of sequences A000225, A033484, and A028399, respectively, in the On-Line Encyclopedia of Integer Sequences [OIES 2010], which notes other combinatorial interpretations of them.

We also conjecture that the tradeoff in Cor. 4.10 generalizes to other protocols.

CONJECTURE 4.12. *There is* no *protocol for* 2ND-PRICE AUCTION$_k$ *that achieves bounded average-case objective PAR (w.r.t. the uniform distribution, as $k \to \infty$) and has sub-exponential communication complexity.*

4.4.1. Subjective privacy for bounded-bisection auctions. We now turn to the subjective privacy of bounded-bisection auctions.

THEOREM 4.13. **(The average-case PAR w.r.t. 1 of the bounded-bisection auction)** The average-case PAR with respect to 1 of the BISECTION AUCTION$_{g(k)}$ is

$$\frac{g(k)+5}{4} - \frac{1}{2^{g(k)+2}} - \frac{1}{2^{k-g(k)+1}} - \frac{g(k)-2}{2^{k+2}}$$

with respect to the uniform distribution.

PROOF. The approach is similar to that in the proof of Thm. 4.9. We start by specializing Eq. 2 to the present case.

Each ideal region in which bidder 1 wins is a rectangle of size 1; each ideal region in which bidder 2 wins is a rectangle of height 1 and width strictly less than 2^k. For a protocol-induced rectangle R, let $j_R = 2^k - |R^I(R)|$. Let $c = g(k)$ and let $i = k - c \geq 0$. Let $T^1_{c,i}$ be the refinement of the 2ND-PRICE-AUCTION$_k$ 1-partition of the k-bit value space induced by the BISECTION-AUCTION$_{g(k)}$. Let $x_{c,i}$ be the number of rectangles in $T^1_{c,i}$ in which bidder 2 (the column player) wins, and let $y_{c,i}$ be the sum, over all rectangles R in which bidder 2 wins, of the quantity $2^{c+i} - |R^I(R)|$. Let $z_{c,i}$ be the number of rectangles R in which bidder 1 (the row player) wins.

Using $\mathsf{PAR}^1_{c,i}$ to denote the PAR w.r.t. bidder 1 in this case (c bisections and $i = k - c$), we may rewrite Eq. 2 as

$$\begin{aligned}\mathsf{PAR}^1_{c,i} &= \frac{1}{2^{2(c+i)}}\left[\left(\sum_{\substack{R^P \text{ in which} \\ 1 \text{ wins}}} |R^I(R^P)|\right) + \left(\sum_{\substack{R^P \text{ in which} \\ 2 \text{ wins}}} |R^I(R^P)|\right)\right] \\ &= \frac{1}{2^{2(c+i)}}\left[(z_{c,i}) + \left(2^{c+i}x_{c,i} - y_{c,i}\right)\right].\end{aligned}$$

We now turn to the computation of $x_{c,i}$, $y_{c,i}$, and $z_{c,i}$.

Following the same approach as in the proof of Thm. 4.9, we have $x_{c+1,i} = 2x_{c,i} + 2^{c+i}$, $y_{c+1,i} = 2y_{c,i} + \sum_{j=1}^{2^{c+i}} j + 2^{c+i}x_{c,i}$, and $z_{c+1,i} = 2z_{c,i} + 2^{2(c+i)}$. With $x_{0,i} = 2^i - 1$, $y_{0,i} =$

$\sum_{j=1}^{2^i-1} j$, and $z_{0,i} = \sum_{j=1}^{2^{c+1}} j$, we obtain

$$\begin{aligned} x_{c,i} &= 2^{c-1}\left(2^i c + 2^{i+1} - 2\right), \\ y_{c,i} &= 2^{c+i-2}\left(2^{c+i}c + 2^{c+i} + 2^i - 2^{c+1} + c\right), \text{ and} \\ z_{c,i} &= 2^{c+i-1}(2^{c+i}+1). \end{aligned}$$

Using these in our expression for $\mathsf{PAR}^1_{c,i}$, we obtain

$$\mathsf{PAR}^1_{c,i} = \frac{c+5}{4} + \frac{2-c}{2^{c+i+2}} - \frac{1}{2^{i+1}} - \frac{1}{2^{c+2}}.$$

Recalling that $k = c + i$ and $g(k) = c$ completes the proof. □

THEOREM 4.14. **(The average-case PAR w.r.t. 2 of the bounded-bisection auction)** The average-case PAR with respect to 2 of the BISECTION AUCTION$_{g(k)}$ is

$$\frac{g(k)+5}{4} - \frac{1}{2^{g(k)+2}} + \frac{g(k)}{2^{k+2}}$$

with respect to the uniform distribution.

PROOF. The approach is essentially the same as in the proof of Thm. 4.13, although the induced partition differs slightly.

Let $c = g(k)$ and let $i = k - c \geq 0$. Let $T^2_{c,i}$ be the refinement of the 2ND-PRICE-AUCTION$_k$ 2-partition of the k-bit value space induced by the BISECTION-AUCTION$_{g(k)}$. Let $u_{c,i}$ be the number of rectangles in $T^2_{c,i}$ in which bidder 1 (the row player) wins, and let $v_{c,i}$ be the sum, over all rectangles R in which bidder 1 wins, of the quantity $2^{c+i} - |R^I(R)|$. Let $w_{c,i}$ be the number of rectangles R in which bidder 2 (the column player) wins. Using $\mathsf{PAR}^2_{c,i}$ to denote the PAR w.r.t. bidder 2 in this case (c bisections and $i = k - c$), we may rewrite Eq. 2 as

$$\mathsf{PAR}^2_{c,i} = \frac{1}{2^{2(c+i)}}\left[\left(2^{c+i}u_{c,i} - v_{c,i}\right) + \left(w_{c,i}\right)\right].$$

Mirroring the approach of the proof of Thm. 4.13, we have $u_{c+1,i} = 2u_{c,i} + 2^{c+i}$, $v_{c+1,i} = 2v_{c,i} + 2^{c+i-1}(2^{c+i}-1) + 2^{c+i}u_{c,i}$, and $w_{c,i} = 2^{c+i-1}(2^{c+i}-1)$. With $u_{0,i} = 2^i$ and $v_{0,i} = 2^{i-1}(2^i - 1)$, we obtain

$$\begin{aligned} u_{c,i} &= 2^{c+i-1}(c+2), \\ v_{c,i} &= 2^{c+i-2}\left(2^{c+i}(c+1) + 2^i - c - 2\right), \text{ and} \\ w_{c,i} &= 2^{c+i-1}(2^{c+i}-1). \end{aligned}$$

Using these in our expression for $\mathsf{PAR}^2_{c,i}$, we obtain

$$\mathsf{PAR}^2_{c,i} = \frac{c+5}{4} - \frac{1}{2^{c+2}} + \frac{c}{2^{c+i+2}}.$$

Recalling that $k = c + i$ and $g(k) = c$ completes the proof. □

Because $g(k) \geq 0$, the average-case PAR with respect to 2 is at least as large as the average-case PAR with respect to 1; this gives the average-case subjective PAR of the BISECTION AUCTION$_{g(k)}$ as follows.

COROLLARY 4.15. **(Average-case subjective PAR of the bounded-bisection auction)** The average-case subjective PAR of the BISECTION AUCTION$_{g(k)}$ is

$$\frac{g(k)+5}{4} - \frac{1}{2^{g(k)+2}} + \frac{g(k)}{2^{k+2}}$$

with respect to the uniform distribution.

5. THE MILLIONAIRES PROBLEM AND PUBLIC GOODS: BOUNDS ON PARS

In this section, we prove upper and lower bounds on the privacy-approximation ratios for two classic problems: Yao's millionaires problem and the provision of a public good.

5.1. Problem Specifications

The millionaires problem.. Two millionaires want to know which one is richer. Each millionaire's wealth is private information known only to him, and the millionaire wishes to keep it that way. The goal is to discover the identity of the richer millionaire while preserving the (subjective) privacy of both parties.

Definition 5.1 (THE MILLIONAIRES PROBLEM$_k$).
Input: $x_1, x_2 \in \{0, \ldots, 2^k - 1\}$ (each represented by a k-bit string)
Output: the identity of the party with the higher value, *i.e.*, $\arg\max_{i \in \{1,2\}} x_i$ (breaking ties lexicographically).

There cannot be a perfectly privacy-preserving communication protocol for THE MILLIONAIRES PROBLEM$_k$ [Kushilevitz 1992]. (Considering any two distinct inputs (i, i) and (j, j), which produce a tie broken in favor of the first millionaire, the reasoning is similar to that used with the example in Fig. 3.) Hence, we are interested in the PARs for this well studied problem.

The public-good problem.. There are two agents, each with a private value in $\{0, \ldots, 2^k - 1\}$ that represents his benefit from the construction of a public project (public good), *e.g.*, a bridge.[2] The goal of the social planner is to build the public project only if the sum of the agents' values is at least its cost c, where, as in [Babaioff et al. 2008], the c is set to be $2^k - 1$.

Definition 5.2 (PUBLIC GOOD$_k$).

Input: $x_1, x_2 \in \{0, \ldots, 2^k - 1\}$ (each represented by a k-bit string)
Output: "Build" if $x_1 + x_2 \geq 2^k - 1$, "Do Not Build" otherwise.

It is easy to show (via Observation 2.13) that for PUBLIC GOOD$_k$, as for THE MILLIONAIRES PROBLEM$_k$, no perfectly privacy-preserving communication protocol exists. Therefore, we are interested in the PARs for this problem.

In Sec. 5.4, we consider a truthful version of this problem. Our results for THE MILLIONAIRES PROBLEM$_k$ and the truthful public-good problem are shown in Table II; the results for the standard public-good problem are essentially those for THE MILLIONAIRES PROBLEM$_k$. Of particular note is the fact that the PAR for the truthful public-good problem is bounded for fixed c (as k grows); unlike most of the problems we consider here, this problem allows good privacy bounds to be realized.

[2]This is a discretization of the classic public good problem, in which the private values are taken from an interval of reals, as in [Blumrosen et al. 2007; Babaioff et al. 2008].

Table II. Average-case PARs for THE MILLIONAIRES PROBLEM$_k$ and TRUTHFUL PUBLIC GOOD$_{k,c}$

Protocol	Average-Case Obj. PAR	Average-Case Subj. PAR
THE MILLIONAIRES PROBLEM$_k$		
All	$\geq 2^k - \frac{1}{2} + 2^{-(k+1)}$	
Bisection Protocol	$\frac{3}{2}2^k - \frac{1}{2}$	$\frac{k}{2} + 1$
TRUTHFUL PUBLIC GOOD$_{k,c}$		
All	$\geq 1 + \frac{c^3}{2^{2k+1}}(1 - \frac{1}{c^2})$	

5.2. The Millionaires Problem

The following theorem shows that not only is perfect subjective privacy unattainable for THE MILLIONAIRES PROBLEM$_k$, but a stronger result holds:

THEOREM 5.3 (A WORST-CASE LOWER BOUND ON SUBJECTIVE PAR). *No communication protocol for* THE MILLIONAIRES PROBLEM$_k$ *has a worst-case subjective PAR less than* $2^{\frac{k}{2}}$.

PROOF. Consider a communication protocol P for THE MILLIONAIRES PROBLEM$_k$. Let R represent the space of possible inputs of millionaire 1, and let C represent the space of possible inputs of millionaire 2. In the beginning, $R = C = \{0, \ldots, 2^k - 1\}$. Consider the first (meaningful) bit q transmitted in course of P's execution. Let us assume that this bit is transmitted by millionaire 1. This bit indicates whether 1's value belongs to one of two disjoint subsets of R, R_1 and R_2, whose union equals R. Because we are interested in the worst case, we can choose adversarially to which of these subsets 1's input belongs. Without loss of generality, let $0 \in R_1$. We decide adversarially that 1's value is in R_1 and set $R = R_1$. Similarly, if q is transmitted by millionaire 2, then we set C to be the subset of C containing 0 in the partition of 2's inputs induced by q. We continue this process recursively for each bit transmitted in P.

Observe that, as long as both R and C contain at least two values, P is incapable of computing THE MILLIONAIRES PROBLEM$_k$. This is because 0 belongs to both R and C, and so P cannot eliminate, for either of the millionaires, the possibility that that millionaire has a value of 0 and the other millionaire has a positive value. Hence, this process will go on until P determines that the value of one of the millionaires is exactly 0, *i.e.*, until either $R = \{0\}$ or $C = \{0\}$. Let us examine these two cases:

— **Case I:** $R = \{0\}$**.** Consider the subcase in which x_2 equals 0. Recall that $0 \in C$, and so this is possible. Observe that, in this case, P determines the exact value of x_1, despite the fact that, in the 2-ideal-monochromatic partition, all 2^k possible values of x_1 are in the same monochromatic rectangle when $x_2 = 0$ (because for all these values 1 wins). Hence, we get a lower bound of 2^k on the subjective privacy-approximation ratio.

— **Case II:** $C = \{0\}$**.** Let m denote the highest input in R. We consider two subcases. If $m \leq 2^{\frac{k}{2}}$, then observe that the worst-case subjective privacy-approximation ratio is at least $2^{\frac{k}{2}}$. In the 2-ideal-monochromatic partition, all 2^k possible values of x_1 are in the same monochromatic rectangle if $x_2 = 0$, and the fact that $m \leq 2^{\frac{k}{2}}$ implies that $|R| \leq 2^{\frac{k}{2}}$.

If, on the other hand, $m > 2^{\frac{k}{2}}$, then consider the case in which $x_1 = m$ and $x_2 = 0$. Observe that, in the 1-ideal-monochromatic partition, all values of millionaire 2 in $\{0, \ldots, m-1\}$ are in the same monochromatic rectangle if $x_1 = m$. However, P will enable millionaire 1 to determine that millionaire 2's value is exactly 0. This implies

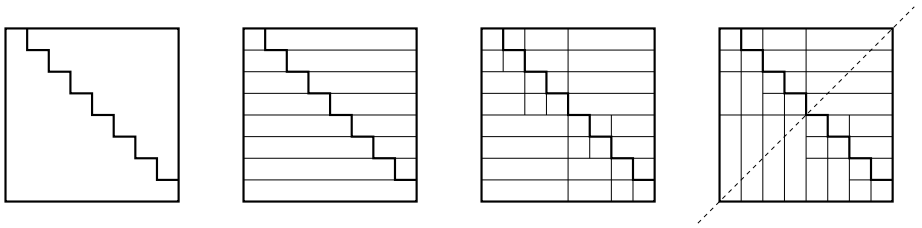

Fig. 6. Left to right: The ideal partition (for $k = 3$) for THE MILLIONAIRES PROBLEM$_k$; the 1-partition of the ideal regions; the 1-induced tiling induced by the BISECTION PROTOCOL; the rearrangement used in the proof of Thm. 5.4

a lower bound of m on the subjective privacy-approximation. We now use the fact that $m > 2^{\frac{k}{2}}$ to conclude the proof.

□

By contrast, we show that fairly good privacy guarantees can be obtained in the average case. We define the BISECTION PROTOCOL for THE MILLIONAIRES PROBLEM$_k$ similarly to the BISECTION AUCTION for 2ND-PRICE AUCTION$_k$: Ask each millionaires whether his value lies in $[0, 2^{k-1})$ or in $[2^{k-1}, 2^k)$; continue this binary search until the millionaires' answers differ, at which point we know which millionaire has the higher value. If the answers never differ the tie is broken in favor of millionaire 1.

We may exactly compute the average-case subjective PAR with respect to the uniform distribution for the BISECTION PROTOCOL applied to THE MILLIONAIRES PROBLEM$_k$. Figure 6 illustrates the approach. The far left of the figure shows the ideal partition (for $k = 3$) of the value space for THE MILLIONAIRES PROBLEM$_k$; these regions are indicated with heavy lines in all parts of the figure. The center-left shows the 1-partition of the regions in the ideal partition; the center-right shows the 1-induced tiling that is induced by the BISECTION PROTOCOL. The far right illustrates how we may rearrange the tiles that partition the bottom-left region in the ideal partition (by reflecting them across the dashed line) to obtain a tiling of the value space that is the same as the tiling induced by applying the BISECTION AUCTION to 2ND-PRICE AUCTION$_k$.

THEOREM 5.4. **The average-case subjective PAR of the bisection protocol** The average-case subjective PAR with respect to the uniform distribution for the BISECTION PROTOCOL applied to THE MILLIONAIRES PROBLEM$_k$ is $\frac{k}{2} + 1$.

PROOF. Given a value of i, consider the i-induced-tiling obtained by running the BISECTION PROTOCOL for THE MILLIONAIRES PROBLEM$_k$ (as in the center-right of Fig. 6 for $i = 1$). Rearrange the rectangles in which player i wins by reflecting them across the line running from the bottom-left corner to the top-right corner (the dashed line in the far right of Fig. 6). This produces a tiling of the value space in which the region in which player 1 wins is tiled by tiles of width 1, and the region in which player 2 wins is tiled by tiles of height 1; in computing the average-case-approximate-privacy with respect to i, the tile-size ratios that we use are the heights (widths) of the tiles to the height (width) of the tile containing all values in that column (row) for which player 1 (2) wins. This tiling and the tile-size ratios in question are exactly as in the computation of the average-case *objective* privacy for 2ND-PRICE AUCTION$_k$; the argument used in Thm. 4.9 (for $g(k) = k$) below completes the proof. □

As with the 2^{nd}-price Vickrey auction, we conjecture that the good PAR that can be achieved with respect to the uniform distribution can be achieved with respect to any distribution. In light of Prop. 5.6, the analogous result for objective PAR does not hold.

CONJECTURE 5.5. *For any probability distribution D over the k-bit input space, the average-case subjective PAR of* THE MILLIONAIRES PROBLEM$_k$ *is linear in k.*

Consider the case in which a third party is observing the interaction of the two millionaires. How much can this observer learn about the private information of the two millionaires? We show that, unlike the case of subjective privacy, good PARs are unattainable even in the average case.

Because the values (i, i) (in which case player 1 wins) and the values $(i, i+1)$ (in which player 2 wins) must all appear in different tiles in any tiling that refines the ideal partition of the value space for THE MILLIONAIRES PROBLEM$_k$, any such tiling must include at least 2^k tiles in which player 1 wins and $2^k - 1$ tiles in which player 2 wins. The total contribution of a tile in which player 1 wins is the number of values in that tile times the ratio of the ideal region containing the tile to the size of the tile, divided by the total number (2^{2k}) of values in the space. Each tile in which player 1 wins thus contributes $\frac{(1+2^k)2^k}{2^{2k+1}}$ to the average-case PAR under the uniform distribution; similarly, each tile in which player 2 wins contributes $\frac{2^k(2^k-1)}{2^{2k+1}}$ to this quantity. This leads directly to the following result.

PROPOSITION 5.6 (A LOWER BOUND ON AVERAGE-CASE OBJECTIVE PAR). *The average-case objective PAR for* THE MILLIONAIRES PROBLEM$_k$ *with respect to the uniform distribution is at least $2^k - \frac{1}{2} + 2^{-(k+1)}$.*

There are numerous different tilings of the value space that achieve this ratio and that can be realized by communication protocols. For the BISECTION PROTOCOL, we obtain the same exponential (in k) growth rate but with a larger constant factor.

PROPOSITION 5.7. **(The average-case objective PAR of the bisection protocol)** The BISECTION PROTOCOL for THE MILLIONAIRES PROBLEM$_k$ obtains an average-case objective PAR of $3 \cdot 2^{k-1} - \frac{1}{2}$ with respect to the uniform distribution.

PROOF. The bisection mechanism induces a tiling that refines the ideal partition and that has $2^{k+1} - 1$ tiles in which the player 1 wins and $2^k - 1$ tiles in which the player 2 wins. The contributions of each of these tiles is as noted above, from which the result follows. □

5.3. The Public-Good Problem

The government is considering the construction of a bridge (a public good) at cost c. Each taxpayer has a k-bit private value that is the utility he would gain from the bridge if it were built. The government wants to build the bridge if and only if the sum of the taxpayers' private values is at least c. In the case that $c = 2^k - 1$, we observe that $\hat{x}_2 = c - x_2$ is again a k-bit value and that $x_1 + x_2 \geq c$ if and only if $x_1 \geq \hat{x}_2$; from the perspective of PAR, this problem is equivalent to solving THE MILLIONAIRES PROBLEM$_k$ on inputs x_1 and $\hat{x}_2$. We may apply our results for THE MILLIONAIRES PROBLEM$_k$ to see that the public-good problem with $c = 2^k - 1$ has exponential average-case objective PAR with respect to the uniform distribution. Section 5.4 discusses average-case objective PAR for a truthful version of the public-good problem.

5.4. Truthful Public-Good Problem

5.4.1. Problem. As in Sec. 5.3, the government is considering the construction of a bridge at cost c. Each taxpayer has a private value that is the utility he would gain

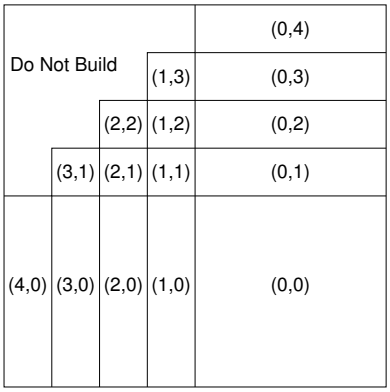

Fig. 7. Ideal partition of the value space for TRUTHFUL PUBLIC GOOD$_{k,c}$ with $k = 3$ and $c = 4$.

from the bridge if it were built, and the government wants to build the bridge if and only if the sum of the taxpayers' private values is at least c. Now, in addition to determining whether to build the bridge, the government incentivizes truthful disclosure of the private values by requiring taxpayer i to pay $c - \sum_{j \neq i} x_j$ if $\sum_{j \neq i} x_j < c$ but $\sum_i x_i \geq c$ (see, *e.g.*, [Nisan 2007] for a discussion of this type of approach). The government should thus learn whether or not to build the bridge and how much, if anything, each taxpayer should pay. The formal description of the function is as follows; the corresponding ideal partition of the value space is shown in Fig. 7, in which regions for which the output is "Build" are just labeled with the appropriate value of (t_1, t_2).

Definition 5.8 (TRUTHFUL PUBLIC GOOD$_{k,c}$).
Input: $c, x_1, x_2 \in \{0, \ldots, 2^k - 1\}$ (each represented by a k-bit string)
Output: "Do Not Build" if $x_1 + x_2 < c$; "Build" and (t_1, t_2) if $x_1 + x_2 \geq c$, where $t_i = c - x_{3-i}$ if $x_{3-i} < c$ and $x_1 + x_2 \geq c$, and $t_i = 0$ otherwise.

5.4.2. Results

PROPOSITION 5.9. **(Average-case objective PAR of** TRUTHFUL PUBLIC GOOD$_{k,c}$**)** The average-case objective PAR of TRUTHFUL PUBLIC GOOD$_{k,c}$ with respect to the uniform distribution is

$$1 + \frac{c^3}{2^{2k+1}}\left(1 - \frac{1}{c^2}\right).$$

PROOF. We may rewrite Eq. 2 as (adding subscripts for the values of k and c in this problem):

$$\mathsf{PAR}_{k,c} = \frac{1}{2^{2k}}\left[\sum_{R_{\mathsf{DNB}}} |R^I(R_{\mathsf{DNB}})| + \sum_{R_{\mathsf{B}}} |R^I(R_{\mathsf{B}})|\right],$$

where the first sum is taken over rectangles R_{DNB} for which the output is "Do Not Build" and the second sum is taken over rectangles R_{B} for which the output is "Build" together with some (t_1, t_2). Using the same argument as for THE MILLIONAIRES PROBLEM$_k$, the first sum must be taken over at least c rectangles; the ideal region containing these rectangles has size $\sum_{i=1}^{c} i = c(c+1)/2$. Considering the second sum, each of the ideal regions containing a protocol-induced rectangle is in fact a rectangle. If the protocol did not further partition these rectangles (and it is easy to see that such

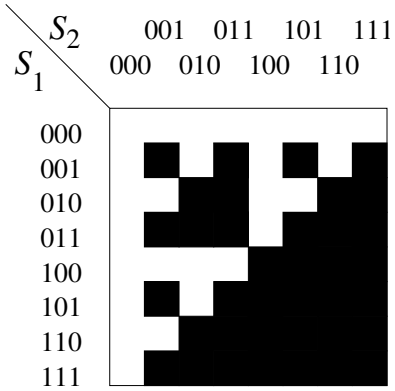

Fig. 8. Ideal monochromatic partition for DISJOINTNESS$_k$ with $k = 3$.

protocols exist) then the total contribution of the second sum is just the total number of inputs for which the output is "Do Not Build" together with some pair (t_1, t_2), *i.e.*, this contribution is $4^k - c(c+1)/2$. We may thus rewrite $\mathsf{PAR}_{k,c}$ as

$$\mathsf{PAR}_{k,c} = \frac{1}{2^{2k}}\left[c\frac{c(c+1)}{2} + 4^k - \frac{c(c+1)}{2}\right] = 1 + \frac{c^3}{2^{2k+1}}(1 - \frac{1}{c^2})$$

□

Unsurprisingly, if we take $c = 2^k - 1$ (as in PUBLIC GOOD$_k$ in Sec. 5.3), we obtain $\mathsf{PAR}_{k,2^k-1} = 2^{k-1} - \frac{1}{2} + \frac{1}{2^k}$, which is essentially half of the average-case PAR for THE MILLIONAIRES PROBLEM$_k$.

6. OVERVIEW OF PROTOCOLS AND RESULTS FOR SET PROBLEMS

We now provide an overview of our PAR results for the set-theoretic problems that we study and discuss their significance. We start with technical definitions of the problems and protocols that we consider here.

6.1. Problems

We define the DISJOINTNESS$_k$ problem as follows:
Problem: DISJOINTNESS$_k$
Input: Sets $S_1, S_2 \subseteq \{1, \ldots, k\}$ encoded by x_1 and x_2.
Output: 1 if $S_1 \cap S_2 = \emptyset$, 0 if $S_1 \cap S_2 \neq \emptyset$.

Figure 8 illustrates the ideal monochromatic partition of the 3-bit value space; inputs for which S_1 and S_2 are disjoint are white, and inputs for which these sets are not disjoint are black.

We define the INTERSECTION$_k$ problem as follows:
Problem: INTERSECTION$_k$
Input: Sets $S_1, S_2 \subseteq \{1, \ldots, k\}$.
Output: The set $S_1 \cap S_2$.

Figure 9 shows the ideal monochromatic partition of the 3-bit value space for INTERSECTION$_k$. The key at the right indicates the output set. (Here, as throughout this paper, we encode $S \subseteq \{1, \ldots, k\}$ as bitstring of length k in which the most significant bit is 1 if $k \in S$, *etc*., so that 1011 encodes $\{1, 2, 4\} \subset \{1, 2, 3, 4\}$; we will abuse notation and identify $x \in \{0, 1\}^k$ with the subset of $\{1, \ldots, k\}$ that it encodes.)

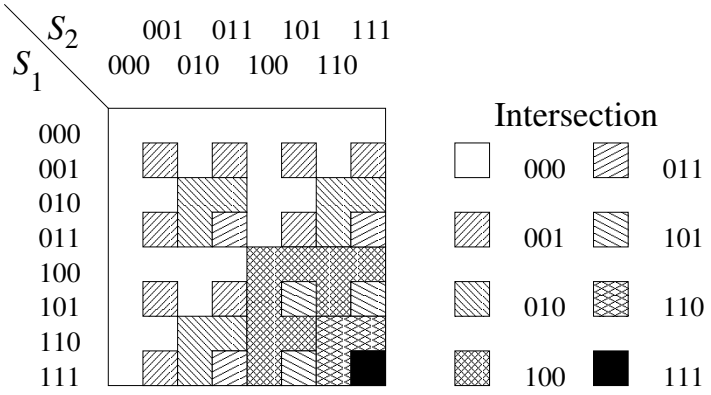

Fig. 9. Ideal monochromatic partition for INTERSECTION$_k$ problem with $k = 3$.

6.2. Protocols

For each problem, we identify three possible protocols for computing the output of the problem. We describe these protocols here; in Sec. 7 we discuss the structure of the tilings that these protocols induce for INTERSECTION$_k$ and illustrate these tilings for $k = 1, 2, 3$.

Trivial protocol. In the trivial protocol, player 1 (w.l.o.g.) sends his input to player 2, who computes the output and sends this back to player 1. This requires the transmission of $k + 1$ bits for DISJOINTNESS$_k$ and $2k$ bits for INTERSECTION$_k$.

1*-first protocol.* In the 1-first protocol, player 1 announces a bit, and player 2 replies with his corresponding bit if its value might affect the output (*i.e.*, if player 1's value for this bit is 1); this continues until the output is determined. In detail, player 1 announces the most significant (first) bit of x_1. After player 1 announces his j^{th} bit, if this bit is 0 and $j < k$, then player 1 announces his $(j+1)^{\text{st}}$ bit. If this bit is 0 and $j = k$, then the protocol terminates (with, if computing DISJOINTNESS$_k$, output 1). If this bit is 1, then player 2 announces the value of his j^{th} bit. If player 2's j^{th} bit is also 1, then for DISJOINTNESS$_k$ the protocol terminates with output 0, and for INTERSECTION$_k$ the protocol continues (with $k + 1 - j$ in the output set); if player 2's bit is 0 and $j < k$, then player 1 announces his $(j + 1)^{\text{st}}$ bit, while if $j = k$, then the protocol terminates.

Alternating protocol. In the alternating protocol, the role of being the first player to announce the value of a particular bit alternates between the players whenever the first player to announce the value of his j^{th} bit announces "0" (in which case the other player does not announce the value of his corresponding bit). This continues until the output is determined. In detail, player 1 starts by announcing the most significant (first) bit of x_1. After player i announces the value of his j^{th} bit, if this bit is 0 and $j < k$, then the other player announces his $j + 1^{\text{st}}$ bit; if i's j^{th} bit is 0 and $j = k$, the protocol terminates (with output 1 if computing DISJOINTNESS$_k$).

If i's j^{th} bit is 1 and the other player had previously announced his j^{th} bit (which would necessarily be 1, else player i would not be announcing his j^{th} bit), then the protocol terminates with output 0 if computing DISJOINTNESS$_k$, or it continues with the other player announcing his $(j+1)^{\text{st}}$ bit (and with $k+1-j$ being part of the output set). If i's j^{th} bit is 1 and the other player had not previously announced his j^{th} bit, then the other player announces his j^{th} bit; if that bit is 0, then player i proceeds as above. If that bit is 1 and DISJOINTNESS$_k$ is being computed, the protocol terminates with

Table III. Summary of average-case results for uniform distribution. Asymptotic results are for $k \to \infty$.

Problem	Protocol	Objective PAR	Subjective PAR	Ratio of Subj. PARs
DISJOINTNESS$_k$	All	$\geq \left(\frac{3}{2}\right)^k$	—	—
	Trivial	$\sim 2^k$	$\sim 2^k$	$\sim 2^k$
	1 First	$\sim 2^k$	$\sim \left(\frac{3}{2}\right)^k$	$\sim \frac{2}{k}\left(\frac{3}{2}\right)^k$
	Alternating	$\sim 2^k$	$\sim \frac{3+2\sqrt{2}}{2}\left(\frac{1+\sqrt{2}}{2}\right)^k$	$\sim \sqrt{2}$
INTERSECTION$_k$	All	$\geq \left(\frac{7}{4}\right)^k$	—	—
	Trivial/1 First	$\left(\frac{7}{4}\right)^k$	$\left(\frac{3}{2}\right)^k$	$\left(\frac{3}{2}\right)^k$
	Alternating	$\left(\frac{7}{4}\right)^k$	$\frac{6}{5}\left(\frac{5}{4}\right)^k$	$\frac{3}{2}$

output 0; if the bit is 1 and INTERSECTION$_k$ is being computed, then player i proceeds as above (and $k+1-j$ will be in the output set).

6.3. Results

Table III summarizes our PAR results for the DISJOINTNESS$_k$ and INTERSECTION$_k$ problems. The rows labeled with "All" describe bounds for all protocols for that problem (as reflected by the inequalities). Asymptotic results are for $k \to \infty$; entries of "—" for bounds on subjective PARs indicate that we do not have results beyond those implied by the PARs for specific protocols. For INTERSECTION$_k$, the results for the trivial and 1-first protocols are shown together; as shown in Lemma 7.1, these protocols induce the same tiling, so the PAR results are the same. All of these results are for average-case objective PARs with respect to the uniform distribution. These include objective and subjective PARs and the ratio of the subjective PARs.

6.3.1. Discussion of results for DISJOINTNESS$_k$. All three protocols have the lowest possible average-case objective PAR for DISJOINTNESS$_k$. They also have average-case subjective PARs that are exponential in k, although the bases differ. When considering these protocols, however, our intuition is that players are much less likely to participate in the trivial and 1-first protocols (if they do so as player 1) than they are to participate in the alternating protocol. This is captured by the comparison of the average-case subjective PAR with respect to the two players in each protocol: In the trivial and 1-first protocols, the subjective PAR with respect to player 2 is exponentially worse than the subjective PAR with respect to player 1; by contrast, in the alternating protocol the subjective PARs differ (asymptotically) by a constant factor. We do not have any absolute lower bound for the average-case subjective PAR for DISJOINTNESS$_k$. However, we conjecture that this grows exponentially.

CONJECTURE 6.1. *The average-case subjective PAR for* DISJOINTNESS$_k$ *with respect to the uniform distribution grows exponentially in* k.

We omit the details of obtaining the results for DISJOINTNESS$_k$ as these use techniques similar to (although requiring greater length) those used for INTERSECTION$_k$.

6.4. Discussion of results for INTERSECTION$_k$

From a high-level perspective, the PAR results for INTERSECTION$_k$ are very similar to those for DISJOINTNESS$_k$. As for their DISJOINTNESS$_k$ variants, all three protocols have exponentially large average-case objective PAR for INTERSECTION$_k$; we show that the average-case objective PAR for INTERSECTION$_k$ is also exponential in k, and we conjecture that this bound can be tightened to match the 2^k asymptotic growth of the average-case objective PAR for all three of these protocols.

CONJECTURE 6.2. *The average-case objective PAR for* INTERSECTION$_k$ *is asymptotic to* 2^k.

All three protocols also have average-case subjective PARs that are exponential in k, although the bases differ. Our intuition that the alternating protocol is significantly better is not captured by the average-case objective and subjective PARs, but we again see it when we consider the ratio of the subjective PARs: In the trivial and 1-first protocols, the subjective PAR for player 1 is exponentially worse than the subjective PAR for player 2; by contrast, in the alternating protocol the subjective PARs differ by a constant factor of $\frac{3}{2}$. In a preliminary version of this paper, we conjectured that the lower bound for the average-case subjective PAR for INTERSECTION$_k$ with respect to the uniform distribution grows exponentially in the number of bits in the input space [Feigenbaum et al. 2010b, Conj. 3.3]. This has since been proved by Ada *et al.*:

THEOREM 6.3 (ADA *et al.*, THM. 3 OF [ADA ET AL. 2012]). *The average-case subjective PAR of* INTERSECTION$_k$ *with respect to the uniform distribution is exponential in* k.

7. PARS FOR INTERSECTION$_K$

7.1. Structure of Protocol-Induced Tilings

First, we observe that for INTERSECTION$_k$, the trivial and 1-first protocols induce the same tiling.

LEMMA 7.1. *The tilings induced by the trivial and* 1*-first protocols for* INTERSECTION$_k$ *are identical.*

PROOF. Given two input pairs (S_1, S_2) and (T_1, T_2), each of these protocols cannot distinguish between the pairs if and only if (1) $S_1 = T_1$ and (2) S_2 and T_2 differ only on elements that are not in $S_1 = T_1$. □

Figure 10 depicts the tilings of the 1-, 2-, and 3-bit value spaces induced by the trivial and 1-first protocols for INTERSECTION$_k$. If we denote by T_k the 1-first-protocol-induced tiling of the k-bit input space, then when we depict T_{k+1} as in Fig. 10, the bottom-left quadrant is $10T_k$ (*i.e.*, the k-bit tiling with 10 prepended to each transcript), each of the top quadrants is $0T_k$, and the bottom-right quadrant is $11T_k$.

Figure 11 depicts the tilings of the 1-, 2-, and 3-bit value spaces induced by the alternating protocol for INTERSECTION$_k$. If we denote by T_k the alternating-protocol-induced tiling of the k-bit value space and depict T_{k+1} as in Fig. 11, the bottom-left quadrant is $10T_k$ (*i.e.*, the k-bit tiling with 10 prepended to each transcript), each of the top quadrants is $0T_k^{\mathsf{T}}$ (*i.e.*, the k-bit tiling reflected across the top-left–bottom-right diagonal), and the bottom-right quadrant is $11T_k$.

7.2. Objective PAR

7.2.1. Lower bound. We obtain the following result for the average-case objective PAR of the INTERSECTION$_k$ problem.

THEOREM 7.2. *The average-case objective PAR of the* INTERSECTION$_k$ *problem with respect to the uniform distribution is* $\left(\frac{7}{4}\right)^k$.

PROOF. We show that $\mathsf{PAR}_{k+1} = \frac{7}{4}\mathsf{PAR}_k$ and that $\mathsf{PAR}_1 = \frac{7}{4}$.

Using Eq. 1, we may write PAR_{k+1} as

$$\mathsf{PAR}_{k+1} = \frac{1}{2^{2(k+1)}} \left(\sum_{R=f^{-1}(0\ldots)} |R^I(R)| + \sum_{R=f^{-1}(1\ldots)} |R^I(R)| \right), \tag{4}$$

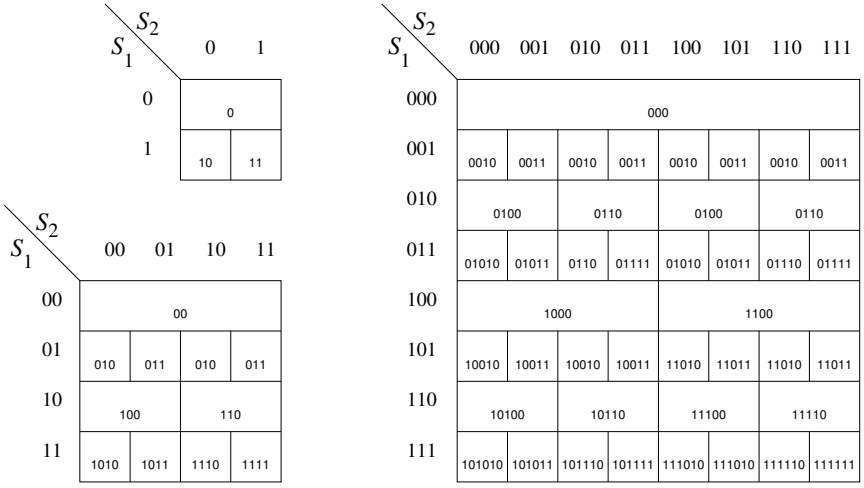

Fig. 10. Partition of the value space for $k = 1$ (top left), 2 (bottom left), and 3 (right) induced by the trivial and 1-first protocols for INTERSECTION$_k$; each rectangle is labeled with the transcript output by the protocol when run on inputs in the rectangle.

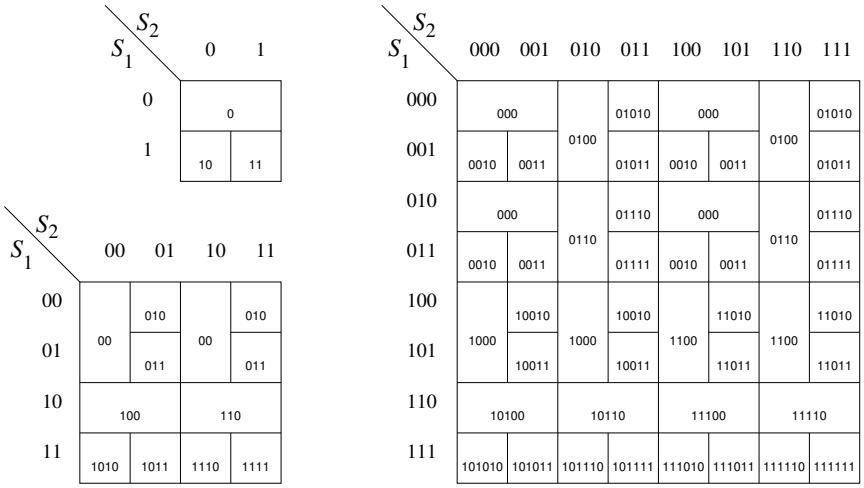

Fig. 11. Partition of the value space for $k = 1$ (top left), 2 (bottom left), and 3 (right) induced by the alternating protocol for INTERSECTION$_k$; each rectangle is labeled with the transcript output by the protocol when run on inputs in the rectangle.

where the first sum is over induced rectangles R in which the intersection set does not contain $k+1$ (*i.e.*, the encoding of the set starts with 0) and the second sum is over induced rectangles R in which the intersection set does contain this element. Observe that the ideal monochromatic partition of the region corresponding to inputs in which $k+1 \in S_1 \cap S_2$ (the bottom-right quadrant when depicted as in Fig. 9) has the same structure as the ideal monochromatic partition of the entire space when only k elements are used. Similarly, the three regions corresponding to $k+1 \notin S_1 \cup S_2$ (top-left quadrant), $k+1 \in S_1 \setminus S_2$ (bottom-left quadrant), and $k+1 \in S_2 \setminus S_1$ (top-right quadrant)

all have this same structure, although each input in these regions belongs to the same monochromatic region as the corresponding inputs in the other two quadrants.

The first observation allows us to rewrite Eq. 4 as

$$\mathsf{PAR}_{k+1} = \frac{1}{4}\left(\frac{1}{2^{2k}}\sum_{R=f^{-1}(0\ldots)}|R^I(R)|\right) + \frac{1}{4}\mathsf{PAR}_k. \tag{5}$$

We now turn to rewriting the term in parentheses.

Consider an input $(0x_1, 0x_2) \in f^{-1}(0x)$ (*i.e.*, $x, x_i \in \{0,1\}^k$ and $x_1 \cap x_2 = x$) in the top-left quadrant of the $(k+1)$-bit input space (when depicted as in Fig. 9). In any monochromatic tiling of this space, $(0x_1, 0x_2)$ may be in the same tile as at most one of the inputs $(0x_1, 1x_2)$ (top-right quadrant) and $(1x_1, 0x_2)$ (bottom-left quadrant)—if both $(0x_1, 1x_2)$ and $(1x_1, 0x_2)$ were in the same tile, then $(1x_1, 1x_2) \in f^{-1}(1x)$ would also be in this tile, violating monochromaticity. If a_x is the minimum number of monochromatic tiles needed to tile the region $f^{-1}(x)$ in the k-bit input space, then at least $2a_x$ monochromatic tiles are needed to tile the region $f^{-1}(0x)$ in the $(k+1)$-bit input space. For any $x \in \{0,1\}^k$, the size of the ideal monochromatic region $f^{-1}(0x)$ is 3 times the size of the monochromatic region $f^{-1}(x)$ in the ideal partition of the input space for k-element sets. Thus the contribution to the sum (for PAR_{k+1}) in Eq. 4 of the rectangles R in $f^{-1}(0x)$ is 6 times the contributions of the contribution to the sum (for PAR_k) of the rectangles R in $f^{-1}(x)$. This allows us to rewrite Eq. 5 as

$$\mathsf{PAR}_{k+1} = \frac{6}{4}\mathsf{PAR}_k + \frac{1}{4}\mathsf{PAR}_k.$$

Finally, the ideal partition for the $\textsc{Intersection}_k$ problem with $k = 1$ requires at least 2 tiles for the region (of size 3) corresponding to an empty intersection and a single tile for the region (of size 1) corresponding to a non-empty intersection. This immediately gives the initial condition $\mathsf{PAR}_1 = \frac{7}{4}$. □

7.2.2. Objective PAR for the trivial and 1-first protocols

PROPOSITION 7.3. *The average-case objective PAR for the trivial and 1-first protocols for the* $\textsc{Intersection}_k$ *problem equals* $\left(\frac{7}{4}\right)^k$.

PROOF. Consider the tiling T_{k+1} of the $(k+1)$-bit value space induced by these protocols. Any tile S in T_k has 3 corresponding tiles in T_{k+1}: the tile whose transcript (in the 1-first protocol) is $10S$, in the bottom-left quadrant; the tile whose transcript is $0S$, which spans the top two quadrants; and the tile whose transcript is $11S$, which is in the bottom-right quadrant. The ideal monochromatic region that contains $0S$ and $10S$ (the same region contains both) in the $(k+1)$-bit value space is 3 times the size of the ideal monochromatic region that contains S in the k-bit value space; the ideal monochromatic region that contains $11S$ is the same size as the ideal monochromatic region that contains S. Thus, we have that $\mathsf{PAR}_{k+1} = \frac{7}{4}\mathsf{PAR}_k$. By inspection, $\mathsf{PAR}_1 = \frac{7}{4}$, finishing the proof. □

7.2.3. Objective PAR for the alternating protocol. Although the recursive tiling structure induced by the alternating protocol is slightly different than that induced by the trivial and 1-first protocols, the argument from the proof of Prop. 7.3 applies essentially unchanged. In particular, even though the structure is different, the tiles in T_{k+1} corresponding to a tile S in T_k are: one tile in the bottom-left quadrant; one tile that spans the top two quadrants; and one tile in the bottom-right quadrant. Thus, we again have $\mathsf{PAR}_{k+1} = \frac{7}{4}\mathsf{PAR}_k$. Again, we also have $\mathsf{PAR}_1 = \frac{7}{4}$, giving us the following proposition.

PROPOSITION 7.4. *The average-case objective PAR for the alternating protocol for the* INTERSECTION$_k$ *problem equals* $\left(\frac{7}{4}\right)^k$. □

7.3. Subjective PAR

7.3.1. Subjective PAR for the trivial and 1-first protocols

PROPOSITION 7.5. *The average-case PAR with respect to player* 1 *of the trivial and* 1*-first protocols for* INTERSECTION$_k$ *is* 1. *The average-case PAR with respect to player* 2 *of the trivial and* 1*-first protocols for* INTERSECTION$_k$*is* $\left(\frac{3}{2}\right)^k$.

PROOF. The 1-partition induced by the trivial protocol is exactly the ideal 1-partition, from which the first claim follows.

For the second claim, we let v_k be the value of the sum in Eq. 1. Let S be a tile in the induced 2-tiling of the k-bit input space; we will also use S to denote the 1-first-protocol transcript that labels S. We now consider the tiles corresponding to S in the induced 2-tiling of the $(k+1)$-bit input space. The tile $10S$ in the bottom-left quadrant is contained in an ideal region that is twice as big as the one that contains S—this ideal region contains points in both the bottom-left and top-left quadrants; the same is true of the tile $0S$ in the top-left quadrant. The tile $0S$ in the top-right quadrant (which is a different 2-induced tile than the one in the top-left quadrant) is contained in an ideal region that is the same size as the ideal region containing S—this ideal region does not contain any points in the bottom-right quadrant. Finally, the tile $11S$ in the bottom-right quadrant is contained in an ideal region that is the same size as the ideal region containing S. Thus, we have that $v_{k+1} = 6v_k$; by inspection, $v_1 = 6$, so $v_k = 6^k$. Note that the average-case PAR with respect to 2 equals $v_k/4^k$, completing the proof. □

COROLLARY 7.6. *The average-case subjective PAR of the trivial and* 1*-first protocols for* INTERSECTION$_k$ *with respect to the uniform distribution is* $\left(\frac{3}{2}\right)^k$.

COROLLARY 7.7. *If* $\mathsf{PAR}_i^{\text{trivial}}$ *denotes the average-case PAR w.r.t.* i *of the trivial protocol for* INTERSECTION$_k$ *w.r.t. the uniform distribution, and if* $\mathsf{PAR}_i^{1-\text{first}}$ *denotes the average-case PAR w.r.t.* i *of the* 1*-first protocol for* INTERSECTION$_k$ *w.r.t. the uniform distribution, then*

$$\frac{\mathsf{PAR}_2^{\text{trivial}}}{\mathsf{PAR}_1^{\text{trivial}}} = \frac{\mathsf{PAR}_2^{1-\text{first}}}{\mathsf{PAR}_1^{1-\text{first}}} = \left(\frac{3}{2}\right)^k.$$

7.3.2. Subjective PAR for the alternating protocol

PROPOSITION 7.8. *The average-case PAR with respect to player* 1 *of the alternating protocol for* INTERSECTION$_k$ *is* $\frac{4}{5}\left(\frac{5}{4}\right)^k$. *The average-case PAR with respect to player* 1 *of the alternating protocol for* INTERSECTION$_k$ *is* $\frac{6}{5}\left(\frac{5}{4}\right)^k$.

PROOF. We let

$$h_k = \sum_S |R_1^I(S)|,$$

where the sum is taken over all induced 1-rectangles ("horizontal rectangles") in the k-bit value space, and we let

$$v_k = \sum_S |R_2^I(S)|,$$

where the sum is taken over all induced 2-rectangles ("vertical rectangles") in the k-bit value space.

Making use of the structure of the tiling, we have that

$$v_{k+1} = 2v_k + 2h_k + h_k + v_k = 3(v_k + h_k),$$

where the summands correspond to the contributions from each quadrant (clockwise from the bottom-left quadrant). We also have

$$h_{k+1} = h_k + 2v_k + h_k = 2(v_k + h_k),$$

where the summands correspond to the contributions from the bottom-left, top-two, and bottom-right quadrants, respectively. By inspection, we have $h_1 = 4$ and $v_1 = 6$; this gives $h_k = 4 \cdot 5^{k-1}$ and $v_k = 6 \cdot 5^{k-1}$. □

As corollaries, we obtain the subjective-PAR results for this protocol that are shown in Table III.

8. DISCUSSION AND FUTURE DIRECTIONS

8.1. Other Notions of Approximate Privacy

By our definitions, the worst-case/average-case PARs of a protocol are determined by the worst-case/expected value of the expression $\frac{|R^I(\mathbf{x})|}{|R^P(\mathbf{x})|}$, where $R^P(\mathbf{x})$ is the monochromatic rectangle induced by P for input $\mathbf{x}$, and $R^I(\mathbf{x})$ is the monochromatic region containing $A(f)_{\mathbf{x}}$ in the ideal monochromatic partition of $A(f)$. That is, informally, we are interested in the ratio of the *size* of the ideal monochromatic region for a specific pair of inputs to the *size* of the monochromatic rectangle induced by the protocol for that pair. More generally, we can define worst-case/average-case PARs with respect to a function g by considering the ratio $\frac{g(R^I(\mathbf{x}),\mathbf{x})}{g(R^P(\mathbf{x}),\mathbf{x})}$. Our definitions of PARs set $g(R, \mathbf{x})$ to be the cardinality of R. This captures the intuitive notion of the indistinguishability of inputs that is natural to consider in the context of privacy preservation. Other definitions of PARs may be appropriate in analyzing other notions of privacy. We suggest a few here; further investigation of these and other definitions provides many interesting avenues for future work.

Probability mass. Given a probability distribution D over the parties' inputs, a seemingly natural choice of g is the probability mass. That is, for any region R, $g(R) = Pr_D(R)$, the probability according to D that the input corresponds to an entry in R. However, a simple example illustrates that this intuitive choice of g is problematic: Consider a problem for which $\{0, \ldots, n\} \times \{i\}$ is a maximal monochromatic region for $0 \leq i \leq n-1$ as illustrated in the left part of Fig. 12. Let P be the communication protocol consisting of a single round in which party 1 reveals whether or not his value is 0; this induces the monochromatic tiling with tiles $\{(0, i)\}$ and $\{(1, i), \ldots, (n, i)\}$ for each i as illustrated in the right part of Fig. 12. Now, for some small $\epsilon > 0$, let D_1 and D_2 be the probability distributions over the inputs $\mathbf{x} = (x_1, x_2)$ such that, for $0 \leq i \leq n-1$ and $1 \leq j \leq n$, $Pr_{D_1}[(x_1, x_2) = (0, i)] = \frac{\epsilon}{n}$, $Pr_{D_1}[(x_1, x_2) = (j, i)] = \frac{1-\epsilon}{n^2}$, $Pr_{D_2}[(x_1, x_2) = (0, i)] = \frac{1-\epsilon}{n}$, and $Pr_{D_2}[(x_1, x_2) = (j, i)] = \frac{\epsilon}{n^2}$. Intuitively, any reasonable definition of PAR should imply that, for D_2, P provides "bad" privacy guarantees (because with high probability it reveals the value of x_1), and, for D_1, P provides "good" privacy (because with high probability it reveals little about x_1). In sharp contrast, choosing g to be the probability mass results in the same average-case PAR in both cases.

One might rightly argue that when the probability distribution is D_2, *i.e.*, when the x_1 is very likely 0, asking party 1 whether $x_1 = 0$ leaks very little additional information beyond what is already implied by D_2. However, information leakage and privacy are two different (though not unrelated) things. A protocol that, with high probability,

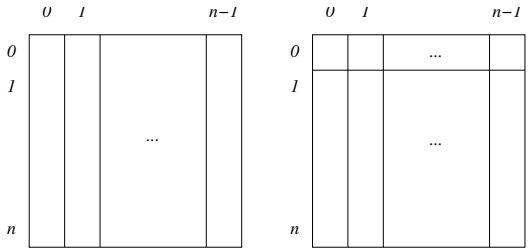

Fig. 12. Maximal monochromatic regions (left) and protocol-induced rectangles (right) for an example showing the deficiencies of PAR definitions based on probability mass.

asks the party for its exact value when that value is not even needed for the computation can hardly be viewed as a protocol with a good PAR, in the same sense that, even if a large fraction of the population has some disease, a protocol that asks someone whether they have the disease to compute an unrelated objective will not be viewed as a protocol with good privacy guarantees.

Information-theoretic approaches. Information-theoretic approaches using conditional entropy are also natural to consider when studying privacy, and these have been used in various settings; such approaches might facilitate the comparison of privacy between different problems. Most relevantly, Bar-Yehuda *et al*. [Bar-Yehuda et al. 2006] defined (in Sec. VII of [Bar-Yehuda et al. 2006]) multiple measures based on the conditional mutual information about one player's value (viewed as a random variable) revealed by the protocol trace and knowledge of the other player's value.

A natural objective (*i.e.*, from the perspective of an outside observer) analogue of the Bar-Yehuda *et al*. approach would use the mutual information $I(\mathsf{X};\mathsf{F})$ between the distribution X on the underlying space and the output of the function F (or $I(\mathsf{X};P)$ for the output of a protocol P). If we want to capture the *effect* of a protocol, a natural quantity to consider is $I(\mathsf{X};P) - I(\mathsf{X};\mathsf{F})$, *i.e.*, how much more information the protocol gives about the underlying distribution than is given by the function; recalling that $I(\mathsf{A};\mathsf{B}) = H(\mathsf{A}) - H(\mathsf{A}|\mathsf{B})$, we see that this is $\mathcal{H}(\mathsf{X}|\mathsf{F}) - \mathcal{H}(\mathsf{X}|P)$ (which has the natural direct interpretation as the decrease in entropy caused by using P). As with the PAR, a larger value of this quantity means that the protocol is less private in some sense.

We now consider two examples that together demonstrate that our definition of PAR is very different than the effect of a function that is captured by $\mathcal{H}(\mathsf{X}|\mathsf{F}) - \mathcal{H}(\mathsf{X}|P)$. In particular, the average-case objective PAR (w.r.t. the uniform distribution) of the first protocol is bounded while this quantity for the second protocol grows linearly in k as $k \to \infty$. By contrast, $\mathcal{H}(\mathsf{X}|\mathsf{F}) - \mathcal{H}(\mathsf{X}|P)$ is slightly larger (for $k \geq 2$) for the first protocol than it is for the second and it approaches similar finite limits for both protocols as $k \to \infty$. Thus, not only do these two approaches seem to disagree on whether the effect of the protocols is asymptotically similar, but they also disagree about *which* protocol has more of a negative effect on the participants' privacy.

The protocols we consider compute the function $f(x_1, x_2) = x_2$, whose induced partition of the 4-bit input space is shown in the left of Fig. 13. In protocol P_1, whose induced partition of the 4-bit input space is illustrated in the center of Fig. 13, each player reveals the most significant bit of his input; if both revealed bits are 0, then each player reveals the second most significant bit of his input. This process is iterated until either one of the revealed bits is 1—at which point player 1 reveals his second bit (if he hasn't already done so) and player 2 reveals the remaining bits of his input—or both players have revealed all of their bits. In protocol P_2, whose induced partition of the 4-bit input space is illustrated in the right of Fig. 13, player 1 reveals his input one

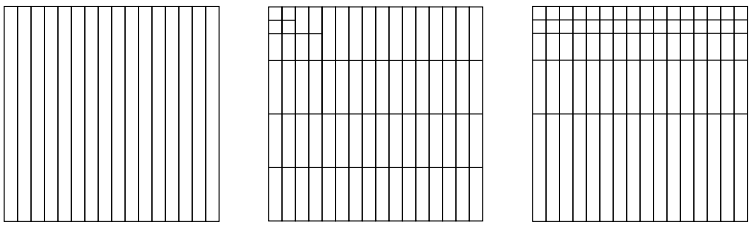

Fig. 13. Induced partitions of the 4-bit value space by the function (left) and the protocols P_1 (center) and P_2 (right).

bit at a time (starting with the most significant bit) until he either reveals a 1 or has revealed his entire input. Player 2 then reveals his entire input.

It is straightforward to compute that the average-case objective PAR (w.r.t. the uniform distribution) of P_1 equals (for $k \geq 2$) $\frac{9}{2} - 2^{-(k-1)}$; for P_2, this quantity equals $k+1$. To compare $\mathcal{H}(\mathsf{X}|\mathsf{F}) - \mathcal{H}(\mathsf{X}|P)$ quantity to PAR, we first observe that

$$\mathcal{H}(\mathsf{X}|\mathsf{F}) - \mathcal{H}(\mathsf{X}|P) = \sum_{x:\mu(x)>0} \mu(x) \log_2 \frac{\mu(R(x))}{\mu(S(x))}, \tag{6}$$

where $R(x)$ and $S(x)$ are the ideal region and protocol-induced rectangle, respectively, that contain the input x. It is then straightforward to see that the quantity in Eq. 6 computed for P_1 equals (for $k \geq 2$) $\frac{25}{12} - \frac{1}{3 \cdot 4^{k-1}}$; for P_2, this information-theoretic quantity equals $2 - 2^{-(k-1)}$.

Other additive functions. In our definition of PAR and in the probability-mass approach, each input $\mathbf{x}$ in a rectangle contributes to $g(R, \mathbf{x})$ in a way that is independent of the other inputs in R. Below, we discuss some natural approaches that violate this condition, but we start by noting that other functions that satisfy this condition may be of interest. For example, taking $g(R, \mathbf{x}) = 1 + \sum_{\mathbf{y} \in R \setminus \mathbf{x}} d(\mathbf{x}, \mathbf{y})$, where d is some distance defined on the input space, gives our original definition of PAR when $d(x, y) = 1 - \delta_{\mathbf{x},\mathbf{y}}$ and might capture other interesting definitions (in which indistinguishable inputs that are farther away from $\mathbf{x}$ contribute more to the privacy for $\mathbf{x}$). (The addition of 1 ensures that the ratio $g(R^I, \mathbf{x})/g(R^P, \mathbf{x})$ is defined, but that can be accomplished in other ways if needed.) Importantly, here and below, the notion of distance that is used might not be a Euclidean metric on the n-player input space $[0, 2^k - 1]^n$. It could instead (and likely would) focus on the problem-specific interpretation of the input space. Of course, there are may possible variations on this (*e.g.*, also accounting for the probability mass).

Maximum distance. We might take the view that a protocol does not reveal much about an input $\mathbf{x}$ if there is another input that is "very different" from $\mathbf{x}$ that the protocol cannot distinguish from $\mathbf{x}$ (even if the total number of things that are indistinguishable from $\mathbf{x}$ under the protocol is relatively small). For some distance d on the input space, we might than take g to be something like $1 + \max_{\mathbf{y} \in R \setminus \{\mathbf{x}\}} d(\mathbf{y}, \mathbf{x})$.

Plausible deniability. One drawback to the maximum-distance approach is that it does not account for the probability associated with inputs that are far from $\mathbf{x}$ (according to a distance d) and that are indistinguishable from $\mathbf{x}$ under the protocol. While there might be an input $\mathbf{y}$ that is far away from $\mathbf{x}$ and indistinguishable from $\mathbf{x}$, the probability of $\mathbf{y}$ might be so small that the observer feels comfortable assuming that $\mathbf{y}$ does not occur. A more realistic approach might be one of "plausible deniability." This makes use of a plausibility threshold—intuitively, the minimum probability that the "far away" inputs(s) (which is/are indistinguishable from

x) must be assigned in order to "distract" the observer from the true input x. This threshold might correspond to, *e.g.*, "reasonable doubt" or other levels of certainty. We then consider how far we can move away from x while still having "enough" mass (*i.e.*, more than the plausibility threshold) associated with the elements indistinguishable from x that are still farther away. We could then take g to be something like $1 + \max\{d_0 | Pr_D(\{\mathbf{y} \in R | d(\mathbf{y}, \mathbf{x}) \geq d_0\})/Pr_D(R) \geq t\}$; other variations might focus on mass that is concentrated in a particular direction from x. (In quantifying privacy, we would expect to only consider those R with positive probability, in which case dividing by $Pr_D(R)$ would not be problematic.) Here we use $Pr_D(R)$ to normalize the weight that is far away from x before comparing it to the threshold t; intuitively, an observer would know that the value is in the same region as x, and so this seems to make the most sense.

Relative rectangle size. One observation is that a bidder likely has a very different view of an auctioneer's being able to tell (when some particular protocol is used) whether his bid lies between 995 and 1005 than he does of the auctioneer's being able to tell whether his bid lies between 5 and 15. In each case, however, the bids in the relevant range are indistinguishable under the protocol from 11 possible bids. In particular, the privacy gained from an input's being distinguishable from a fixed number of other inputs may (or may not) depend on the context of the problem and the intended interpretation of the values in the input space. This might lead to a choice of g such as $diam_d(R)/|\mathbf{x}|$, where $diam_d$ is the diameter of R with respect to some distance d and $|\mathbf{x}|$ is some (problem-specific) measure of the size of x (*e.g.*, bid value in an auction). Numerous variations on this are natural and may be worth investigating.

8.2. Open Questions

There are many interesting directions for future research:

— As discussed in the previous subsection, the definition and exploration of other notions of PARs is a challenging and intriguing direction for future work.

— We have shown that, for both 2ND-PRICE AUCTION$_k$ and THE MILLIONAIRES PROBLEM$_k$, reasonable average-case PARs with respect to the uniform distribution are achievable. We conjecture that our upper bounds for these problems extend to *all* possible distributions over inputs.

— An interesting open question is proving lower bounds on the average-case PARs for 2ND-PRICE AUCTION$_k$ and THE MILLIONAIRES PROBLEM$_k$.

— Lower bounds on the average-case subjective PARs for DISJOINTNESS$_k$ and INTERSECTION$_k$ would be interesting; as noted above, we conjecture that these are exponential in k.

— It would be interesting to apply the PAR framework presented in this paper to other functions.

— The extension of our PAR framework to the n-party communication model is a challenging direction for future research.

— Starting from the same place that we did, namely [Chor and Kushilevitz 1991; Kushilevitz 1992], Bar-Yehuda *et al.* [Bar-Yehuda et al. 2006] provided three definitions of approximate privacy. The one that seems most relevant to the study of privacy-approximation ratios is their notion of *h-privacy*. It would be interesting to know exactly when and how it is possible to express PARs in terms of h-privacy and *vice versa*.

ACKNOWLEDGMENTS

We are grateful to audiences at *ACM EC'10*, University Residential Centre of Bertinoro, Boston University, DIMACS, the University of Massachusetts, Northwestern, Princeton, and Rutgers for helpful questions and feedback. We thank Omri Weinstein for helpful discussions, and we thank the referees for helpful feedback.

REFERENCES

Anil Ada, Arkadev Chattopadhyay, Stephen Cook, Lila Fontes, Michal Koucky, and Toniann Pitassi. 2012. The Hardness of Being Private. In *Proceedings of the 2012 IEEE Conference on Computational Complexity (CCC) (CCC '12)*. IEEE Computer Society, Washington, DC, USA, 192–202. DOI:http://dx.doi.org/10.1109/CCC.2012.24

Moshe Babaioff, Liad Blumrosen, Moni Naor, and Michael Schapira. 2008. Informational Overhead of Incentive Compatibility. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC '08)*. ACM, New York, NY, USA, 88–97. DOI:http://dx.doi.org/10.1145/1386790.1386807

R. Bar-Yehuda, B. Chor, E. Kushilevitz, and A. Orlitsky. 2006. Privacy, Additional Information and Communication. *IEEE Trans. Inf. Theor.* 39, 6 (Sept. 2006), 1930–1943. DOI:http://dx.doi.org/10.1109/18.265501

Amos Beimel, Paz Carmi, Kobbi Nissim, and Enav Weinreb. 2008. Private Approximation of Search Problems. *SIAM J. Comput.* 38, 5 (Dec. 2008), 1728–1760. DOI:http://dx.doi.org/10.1137/060671899

Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. 1988. Completeness Theorems for Non-cryptographic Fault-tolerant Distributed Computation. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing (STOC '88)*. ACM, New York, NY, USA, 1–10. DOI:http://dx.doi.org/10.1145/62212.62213

Liad Blumrosen, Noam Nisan, and Ilya Segal. 2007. Auctions with Severely Bounded Communication. *J. Artif. Int. Res.* 28, 1 (March 2007), 233–266. DOI:http://dx.doi.org/10.1613/jair.2081

Felix Brandt and Tuomas Sandholm. 2008. On the Existence of Unconditionally Privacy-Preserving Auction Protocols. *ACM Trans. Inf. Syst. Secur.* 11, 2, Article 6 (May 2008), 21 pages. DOI:http://dx.doi.org/10.1145/1330332.1330338

David Chaum, Claude Crépeau, and Ivan Damgard. 1988. Multiparty Unconditionally Secure Protocols. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing (STOC '88)*. ACM, New York, NY, USA, 11–19. DOI:http://dx.doi.org/10.1145/62212.62214

Benny Chor and Eyal Kushilevitz. 1991. A Zero-one Law for Boolean Privacy. *SIAM J. Discret. Math.* 4, 1 (Jan. 1991), 36–47. DOI:http://dx.doi.org/10.1137/0404004

Marco Comi, Bhaskar Dasgupta, Michael Schapira, and Venkatakumar Srinivasan. 2012. On Communication Protocols That Compute Almost Privately. *Theor. Comput. Sci.* 457 (Oct. 2012), 45–58. DOI:http://dx.doi.org/10.1016/j.tcs.2012.07.008

Yevgeniy Dodis, Shai Halevi, and Tal Rabin. 2000. A Cryptographic Solution to a Game Theoretic Problem. In *Proceedings of the 20th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '00)*. Springer-Verlag, London, UK, UK, 112–130. DOI:http://dx.doi.org/10.1007/3-540-44598-6_7

Cynthia Dwork. 2006. Differential Privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II (ICALP'06)*. Springer-Verlag, Berlin, Heidelberg, 1–12. DOI:http://dx.doi.org/10.1007/11787006_1

Joan Feigenbaum, Yuval Ishai, Tal Malkin, Kobbi Nissim, Martin J. Strauss, and Rebecca N. Wright. 2006. Secure Multiparty Computation of Approximations. *ACM Trans. Algorithms* 2, 3 (July 2006), 435–472. DOI:http://dx.doi.org/10.1145/1159892.1159900

Joan Feigenbaum, Aaron D. Jaggard, and Michael Schapira. 2010a. Approximate Privacy: Foundations and Quantification (Extended Abstract). In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC '10)*. ACM, New York, NY, USA, 167–178. DOI:http://dx.doi.org/10.1145/1807342.1807369

Joan Feigenbaum, Aaron D. Jaggard, and Michael Schapira. 2010b. *Approximate privacy: PARs for set problems*. Technical Report 2010-01. DIMACS. Also arXiv:1001.3388.

Joan Feigenbaum and Scott Shenker. 2002. Distributed Algorithmic Mechanism Design: Recent Results and Future Directions. In *Proceedings of the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM '02)*. ACM, New York, NY, USA, 1–13. DOI:http://dx.doi.org/10.1145/570810.570812

Yuzo Fujishima, David McAdams, and Yoav Shoham. 1999. Speeding Up Ascending-bid Auctions. In *Proceedings of the 16th International Joint Conference on Artifical Intelligence - Volume 1 (IJCAI'99)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 554–559.

Arpita Ghosh and Aaron Roth. 2011. Selling Privacy at Auction. In *Proceedings of the 12th ACM Conference on Electronic Commerce (EC '11)*. ACM, New York, NY, USA, 199–208. DOI:http://dx.doi.org/10.1145/1993574.1993605

Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. 2009. Universally Utility-maximizing Privacy Mechanisms. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC '09)*. ACM, New York, NY, USA, 351–360. DOI:http://dx.doi.org/10.1145/1536414.1536464

Elena Grigorieva, P. Jean-Jacques Herings, Rudolf MüLler, and Dries Vermeulen. 2006. The Communication Complexity of Private Value Single-item Auctions. *Oper. Res. Lett.* 34, 5 (Sept. 2006), 491–498. DOI:http://dx.doi.org/10.1016/j.orl.2005.07.011

Elena Grigorieva, P. Jean-Jacques. Herings, Rudolf Mller, and Dries Vermeulen. 2007. The private value single item bisection auction. *Economic Theory* 30, 1 (2007), 107–118. DOI:http://dx.doi.org/10.1007/s00199-005-0032-z

Elena Grigorieva, P. Jean-Jacques Herings, Rudolf Mller, and Dries Vermeulen. 2010. On the Fastest Vickrey Algorithm. *Algorithmica* 58, 3 (2010), 566–590. DOI:http://dx.doi.org/10.1007/s00453-009-9285-4

Shai Halevi, Robert Krauthgamer, Eyal Kushilevitz, and Kobbi Nissim. 2001. Private Approximation of NP-hard Functions. In *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing (STOC '01)*. ACM, New York, NY, USA, 550–559. DOI:http://dx.doi.org/10.1145/380752.380850

P. Jean-Jacques Herings, Rudolf Müller, and Dries Vermeulen. 2009. Bisection Auctions. *SIGecom Exch.* 8, 1, Article 6 (July 2009), 5 pages. DOI:http://dx.doi.org/10.1145/1598780.1598786

Eyal Kushilevitz. 1992. Privacy and Communication Complexity. *SIAM J. Discret. Math.* 5, 2 (May 1992), 273–284. DOI:http://dx.doi.org/10.1137/0405021

Eyal Kushilevitz and Noam Nisan. 1997. *Communication Complexity*. Cambridge University Press, New York.

Andrew McGregor, Ilya Mironov, Toniann Pitassi, Omer Reingold, Kunal Talwar, and Salil Vadhan. 2010. The Limits of Two-Party Differential Privacy. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS '10)*. IEEE Computer Society, Washington, DC, USA, 81–90. DOI:http://dx.doi.org/10.1109/FOCS.2010.14

S. Muthukrishnan. 2009. Challenges in Designing Ad Exchanges. (September 2009). Talk at the *NSF Workshop on Research Issues at the Interface of Computer Science and Economics*, Cornell University.

Moni Naor, Benny Pinkas, and Reuban Sumner. 1999. Privacy Preserving Auctions and Mechanism Design. In *Proceedings of the 1st ACM Conference on Electronic Commerce (EC '99)*. ACM, New York, NY, USA, 129–139. DOI:http://dx.doi.org/10.1145/336992.337028

Noam Nisan. 2007. Introduction to Mechanism Design (for Computer Scientists). In *Algorithmic Game Theory*, Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani (Eds.). Cambridge University Press, New York, Chapter 9, 209–242.

Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky. 2012. Privacy-aware Mechanism Design. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC '12)*. ACM, New York, NY, USA, 774–789. DOI:http://dx.doi.org/10.1145/2229012.2229073

OIES. 2010. The On-Line Encyclopedia of Integer Sequences. Published electronically at http://oeis.org. (2010).

Xin Sui and Craig Boutilier. 2011. Efficiency and Privacy Tradeoffs in Mechanism Design. In *AAAI Conference on Artificial Intelligence*. http://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3527

William Vickrey. 1961. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance* 16, 1 (1961), 8–37. DOI:http://dx.doi.org/10.1111/j.1540-6261.1961.tb02789.x

Andrew C. Yao. 1982. Protocols for Secure Computations. In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (SFCS '82)*. IEEE Computer Society, Washington, DC, USA, 160–164. DOI:http://dx.doi.org/10.1109/SFCS.1982.88

Andrew Chi-Chih Yao. 1979. Some Complexity Questions Related to Distributive Computing (Preliminary Report). In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing (STOC '79)*. ACM, New York, NY, USA, 209–213. DOI:http://dx.doi.org/10.1145/800135.804414

Andrew Chi-Chih Yao. 1986. How to Generate and Exchange Secrets. In *Proceedings of the 27th Annual Symposium on Foundations of Computer Science (SFCS '86)*. IEEE Computer Society, Washington, DC, USA, 162–167. DOI:http://dx.doi.org/10.1109/SFCS.1986.25

Received M YYYY; revised M YYYY; accepted M YYYY