DIMACS Technical Report 2010-01
Version of 2010–01–19

Approximate Privacy: PARs for Set Problems

by

Joan Feigenbaum[1,2]
Department of Computer Science
Yale University
joan.feigenbaum@yale.edu

Aaron D. Jaggard[3,4]
DIMACS
Rutgers University
adj@dimacs.rutgers.edu

Michael Schapira[5]
Department of Computer Science
Yale University
and
Computer Science Division
University of California, Berkeley
michael.schapira@yale.edu

[1]DIMACS Permanent Member
[2]Supported in part by NSF grants 0331548 and 0716223 and IARPA grant FA8750-07-0031.
[3]DIMACS Assistant Research Professor and Permanent Member
[4]Supported in part by NSF grants 0751674 and 0753492.
[5]Supported by NSF grant 0331548.

ABSTRACT

In previous work (DIMACS TR 2009-14), we introduced the Privacy Approximation Ratio (PAR) and used it to study the privacy of protocols for second-price Vickrey auctions and Yao's millionaires problem. Here, we study the PARs of multiple protocols for both the disjointness problem (in which two participants, each with a private subset of $\{1, \ldots, k\}$, determine whether their sets are disjoint) and the intersection problem (in which the two participants, each with a private subset of $\{1, \ldots, k\}$, determine the intersection of their private sets).

We show that the privacy, as measured by the PAR, provided by any protocol for each of these problems is necessarily exponential (in k). We also consider the ratio between the subjective PARs with respect to each player in order to show that one protocol for each of these problems is significantly fairer than the others (in the sense that it has a similarly bad effect on the privacy of both players).

1 Introduction

Widespread use of computers and networks in almost all aspects of daily life has led to a proliferation of sensitive electronic data records and thence to extensive study of privacy-preserving computation. One fruitful approach is based on the combinatorial characterization of privately computable functions put forth by Chor and Kushilevitz [4] and the subsequent communication-complexity analysis of privately computable functions by Kushilevitz [11]. Using this approach, one can show, for example, that Yao's millionaires' problem [15] is not perfectly privately computable [4] and that the two-bidder, 2^{nd}-price Vickrey auction is perfectly privately computable but only at the cost of and exponential amount of communication by the bidders [3].

Motivated by the fact that functions of interest may not be perfectly privately computable or may be so only by impractically costly protocols, we began in [8] a communication-complexity-based investigation of approximate privacy. We formulated both worst-case and average-case versions of the *privacy-approximation ratio* (PAR) of a function f in order to quantify the *amount* of privacy that can be preserved by a protocol that computes f and studied the tradeoff between approximate privacy and communication complexity in protocols for the millionaires' problem and the two-bidder, 2^{nd}-price Vickrey auction.

Informally, 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. (This notion can be extended naturally to protocols involving more than two participants, but we do not consider the more general notion in this paper.) Chor and Kushilevitz [4, 11] formalize this notion of privacy using the communication-complexity-theoretic notions of the *ideal monochromatic regions* of a function f and the *monochromatic rectangles* of a protocol P that computes f. Every two-input function f can be represented by a two-dimensional matrix $A(f)$ in which $A(f)_{(x_1,x_2)} = f(x_1, x_2)$. In the partition of $A(f)$ into the ideal monochromatic regions of f, the entries $A(f)_{(x_1,x_2)}$ and $A(f)_{(y_1,y_2)}$ are in the same region if and only if $f(x_1, x_2) = f(y_1, y_2)$; if f is perfectly privately computable, then there is a protocol P for f that partitions $A(f)$ into a set of monochromatic rectangles that is exactly equal to the set of ideal monochromatic regions of f. For functions that are not perfectly privately computable, our notions of approximate privacy [8] quantify the worst-case and average-case ratios between the size of an ideal monochromatic region of f and the corresponding monochromatic rectangle in the partition induced by a maximally privacy-preserving protocol for f.

In this paper, we apply our PAR framework to 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 to 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.

1.1 Our Findings

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. 2.3) and argue that this captures some intuitive sense of fairness. Table 1 in Sec. 3 summarizes our results for PAR values for the various problems and protocols that we consider here; the corresponding theorems and proofs are in Secs. 4 and 5.

1.2 Related Work: Defining Privacy-Preserving Computation

In addition to Brandt and Sandholm [3], who used Kushilevitz's formulation of privacy-preserving computation to prove an exponential lower bound on the communication complexity of privacy-preserving 2^{nd}-price Vickrey auctions, the privacy work of Bar-Yehuda *et al.* [1] is also based on the communication-complexity framework of [4, 11].

Among other approaches to privacy-preserving computation, the most extensively developed is that of *secure, multiparty computation* (SMC). As observed by Brandt and Sandholm [3], bidders' privacy in online auctions, which was our original motivation as well as theirs, could in principle be achieved by starting with a strategyproof mechanism and then having the agents themselves compute the outcome and payments using an SMC protocol. This approach has been followed successfully by, for example, Dodis, Halevi, and Rabin [5] and Naor, Pinkas, and Sumner [14] but, as discussed in more detail [3, 8], can in general require assumptions about the strategic nature of the computational nodes that do not apply to bidders in auctions, unproven cryptographic assumptions, or excessive communication costs. Thus, non-SMC approaches are worth pursuing.

In our study of PAR, we consider protocols that compute exact results but preserve privacy only approximately. Several works, including [2, 7, 10], have considered protocols that compute approximate results in a privacy-preserving manner, but they are unrelated to the questions we ask here. Similarly, definitions and techniques from *differential privacy* [6] (and its mechanism-design extensions [9, 13]) are aimed at computing approximate results and are inapplicable to the problems that we study here.

1.3 Paper Outline

In Sec. 2, we review the PAR framework of [8] and discuss the ratios of average-case subjective PARs. Section 3 gives formal definitions of the problems we study, describes the protocols for these problems that we consider, and gives a summary and discussion of our PAR results. Sections 4 and 5 give the full statements and proofs of our PAR results. Section 6 discusses avenues for future work. Appendix A provides additional background about our approach. Sections 2.1, 2.2, and App. A are drawn from [8]; we include them here for the convenience of the reader.

2 Privacy Approximation Ratios

We now review our formulations of Privacy Approximation Ratios (PARs) [8]. We refer readers to Section A.2 of the Appendix below for a more thorough explanation. We assume that the reader is familiar with Yao's model of two-party communication. Readers unfamiliar with this material should refer to Section A.1 of the Appendix below or, for a more in-depth treatment, to Kushilevitz and Nisan [12].

Chor and Kushilevitz [4,11] put forth definitions and characterizations of perfectly private communication protocols. Their framework was further developed in [8], where we introduced the notion of PARs. In this paper, as in [8], we deal only with *deterministic* communication protocols, but the framework can be extended to randomized protocols.

As explained in the previous section, there are natural problems for which perfect privacy is either impossible or very costly (in terms of communication complexity) to obtain. *Privacy-approximation ratios* (PARs) allow us to quantify how well a protocol preserves privacy relative to the ideal (but perhaps impossible to implement) computation of the outcome of a problem. Approximate privacy has both worst-case and average-case formulations.

2.1 Worst-Case PARs

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 $A(f)$.

For any communication protocol P for a function f, let $R^P(x_1, x_2)$ denote the monochromatic rectangle in $A(f)$ induced by P for the pair of inputs (x_1, x_2). Let $R^I(x_1, x_2)$ denote the maximal monochromatic region in $A(f)$ containing $A(f)_{(x_1,x_2)}$, *i.e.*, the maximal set of entries in $A(f)$ that contain the value $f(x_1, x_2)$. Intuitively, $R^P(x_1, x_2)$ is the set of inputs that are indistinguishable from (x_1, x_2) to this particular protocol P. Similarly, $R^I(x_1, x_2)$ is the set of inputs that *would be* indistinguishable from (x_1, x_2) to a perfectly private protocol if such a protocol existed. We wish to quantify how far P is from a hypothetical ideal protocol in terms of indistinguishability of inputs. Let $|R|$ denote the *size* or cardinality of R, *i.e.*, the number of inputs in R.

Definition 2.1 (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.*

Given any region R in the matrix $A(f)$, if party 1's private input is x, then party 1 can use this knowledge to eliminate all entries in R outside of row x; similarly, party 2 can eliminate all parts of R outside of the appropriate column. Hence, the other parties should be concerned not with all of R but rather with what we call the *i-partitions* of R.

Definition 2.2 (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.

Definition 2.3 (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 2.4 (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.

If P is a communication protocol for the function f, then we let $R_i^P(x_1,x_2)$ denote the monochromatic rectangle containing $A(f)_{(x_1,x_2)}$ in the i-induced tiling for P. Similarly, we let $R_i^I(x_1,x_2)$ denote the monochromatic rectangle containing $A(f)_{(x_1,x_2)}$ in the i-ideal monochromatic partition of $A(f)$.

Definition 2.5 (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 2.6 (Worst-case subjective PAR of P)**.** The *worst-case subjective privacy-approximation ratio* of communication protocol P for function f is the maximum, over $i = 1, 2$, of the worst-case privacy-approximation ratio with respect party i.

Definition 2.7 (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.

2.2 Average-Case PARs

As we showed in [8], good approximate privacy may be just as unobtainable as perfect privacy if one insists on worst-case bounds. Thus, we also consider average-case PAR, *i.e.*, the *average* ratio between the size of the monochromatic rectangle containing the private inputs and the corresponding region in the ideal monochromatic partition.

Definition 2.8 (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

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

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 i of the average-case PAR with respect to player i. Finally, 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 a probability-mass-based definition of average-case PAR should be used instead, *i.e.*, that the occurrences of set cardinality in the quantity considered in Def. 2.8 should be replaced by the probability measure of the regions in question. However, as we discuss in [8], 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.

2.3 Ratios of Subjective PARs

Here we introduce a new quantity that we did not consider in [8]. 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) subjective PARs* to be

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

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.

3 Overview of Problems, Results, and Protocols

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

3.1 Problems

We define the $\textsc{Disjointness}_k$ problem as follows:
Problem: $\textsc{Disjointness}_k$
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 1 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.

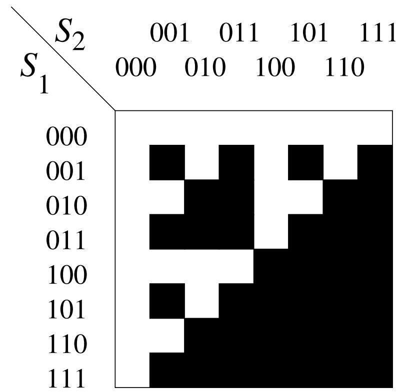

Figure 1: Ideal monochromatic partition for DISJOINTNESS_k with $k = 3$.

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 2 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.)

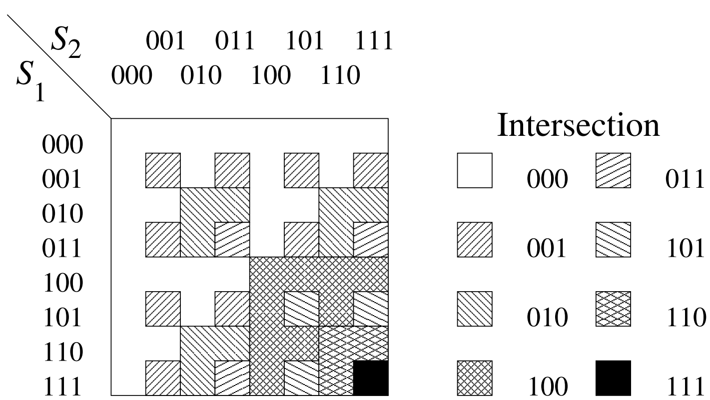

Figure 2: Ideal monochromatic partition for INTERSECTION_k problem with $k = 3$.

3.2 Protocols

For each problem, we identify three possible protocols for computing the output of the problem. We describe these protocols here; in Secs. 4 and 5 we discuss the structure of the tilings that these protocols induce 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 determines 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 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).

3.3 Results

Table 1 summarizes our PAR results for the $\textsc{Disjointness}_k$ and $\textsc{Intersection}_k$ 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 $\textsc{Intersection}_k$, the results for the trivial and 1-first protocols are shown together; as shown in Lemma 5.1, 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.

Problem	Protocol	Objective PAR	Subjective PAR	Ratio of Subj. PARs
$\textsc{Disjointness}_k$	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	$\sim 2^k$	$\sim \frac{3+2\sqrt{2}}{2}\left(\frac{1+\sqrt{2}}{2}\right)^k$	$\sim \sqrt{2}$
$\textsc{Intersection}_k$	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	$\left(\frac{7}{4}\right)^k$	$\frac{6}{5}\left(\frac{5}{4}\right)^k$	$\frac{3}{2}$

Table 1: Summary of results. Asymptotic results are for $k \to \infty$.

3.3.1 Discussion of results for $\textsc{Disjointness}_k$

All three protocols have the lowest possible average-case objective PAR for $\textsc{Disjointness}_k$. They also have average-case subjective PARs that are exponential in k, although the bases differ. When considering these protocols (and the tilings they induce as depicted in Sec. 4), 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 $\textsc{Disjointness}_k$. However, we conjecture that this grows exponentially.

Conjecture 3.1. *The average-case subjective PAR for* $\textsc{Disjointness}_k$ *with respect to the uniform distribution grows exponentially in* k.

3.4 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 3.2. *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 $\frac{3}{2}$. We do not have any absolute lower bound for the average-case subjective PAR for INTERSECTION$_k$. However, as for DISJOINTNESS$_k$, we conjecture that this grows exponentially.

Conjecture 3.3. *The average-case subjective PAR for* INTERSECTION$_k$ *with respect to the uniform distribution grows exponentially in* k*.*

4 PARs for DISJOINTNESS$_k$

4.1 Structure of Protocol-Induced Tilings

The tiling induced by the trivial protocol is straightforward. For every input $S_1 \neq 0^k$ held by player 1, there are two monochromatic rectangles in the corresponding row of the input space: $\{(S_1, S_2)|S_2 \cap S_1 \neq \emptyset\}$ and $\{(S_1, S_2)|S_2 \cap S_1 = \emptyset\}$. The row corresponding to $S_1 = 0^k$ forms a single monochromatic rectangle.

Figure 3 depicts the 1-first-protocol-induced tiling of the 1-, 2-, and 3-bit input spaces. Each tile is labeled with the transcript produced by the protocol on inputs from that tile; note that some tiles are depicted as non-contiguous regions. When the input space is depicted as in Fig. 3 (*i.e.*, with the possible values of S_1 and S_2 arranged in increasing lexicographic order from the top-left corner), the tiling of the $k+1$-bit input space induced by the 1-first protocol can be obtained as follows. Let T_k be the 1-first-protocol-induced tiling of the k-bit input space. The top-left and top-right quadrants of T_{k+1} are copies of T_k; in each of these quadrants, a trace in T_{k+1} is the corresponding trace in T_k prepended with 0. The bottom-left quadrant of T_{k+1} is another copy of T_k, with each trace in this part of T_{k+1} being obtained by prepending 10 to the corresponding trace in T_k. The bottom-right quadrant of T_{k+1} is a single rectangle whose trace is 11.

Figure 4 shows the partition of the 1-, 2-, and 3-bit input spaces induced by the alternating protocol; each induced rectangle is labeled with the corresponding transcript (note that some

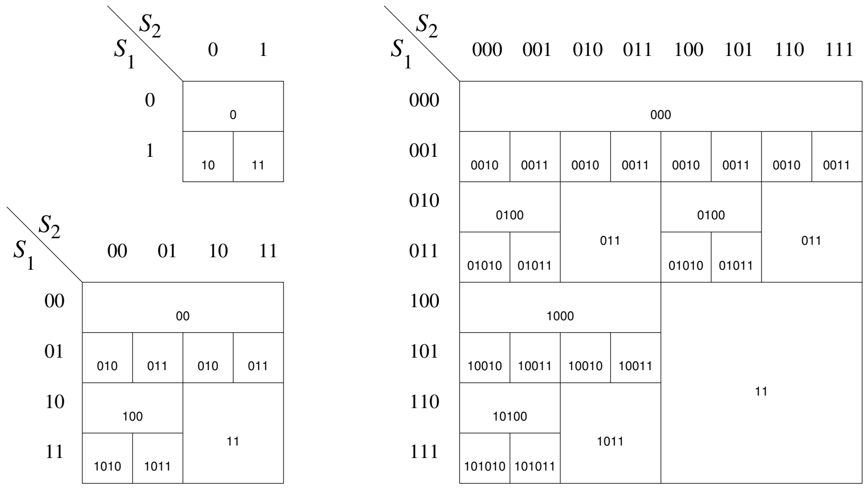

Figure 3: Partition of the value space for $k = 1$ (top left), 2 (bottom left), and 3 (right) induced by the 1-first protocol for DISJOINTNESS$_k$; each rectangle is labeled with the transcript output by the protocol when run on inputs in the rectangle.

rectangles appear as non-contiguous regions in the figure). If we denote by T_k the tiling of the k-bit space induced by the alternating protocol as depicted in Fig. 4, then the bottom-left quadrant of T_{k+1} has the same structure as T_k, with the transcript for a tile in T_{k+1} obtained by prepending 10 to the transcript for the corresponding tile in T_k. Each of the top quadrants has the same structure as the reflection of T_k across the top-left-to-bottom-right diagonal; the corresponding rectangles in these quadrants actually form single rectangles, and the associated transcript is obtained by prepending 0 to the transcript for the corresponding rectangle in T_k. Finally, the bottom-right quadrant is a single rectangle that always has the transcript 11.

4.2 Objective PAR

4.2.1 Objective PAR for the DISJOINTNESS$_k$ problem

Lemma 4.1. *In the ideal partition induced by* DISJOINTNESS$_k$*, at least* 2^k *rectangles are required to tile the region* $f^{-1}(1)$*.*

Proof. As shown in, *e.g.*, [12], the 2^k input pairs $(S, \{1, \ldots, k\} \setminus S)$ form a "fooling set"—no two of these input pairs can belong to the same monochromatic rectangle. □

Corollary 4.2. *The average-case objective PAR of* DISJOINTNESS$_k$ *with respect to the uniform distribution is at least* $\left(\frac{3}{2}\right)^k$*.*

Proof. The contribution to the sum in Eq. 1 from the protocol-induced tiles $S \subset f^{-1}(1)$ must be at least $2^k \cdot 3^k$, so the average-case objective PAR with respect to the uniform distribution is at least $\left(\frac{3}{2}\right)^k$. □

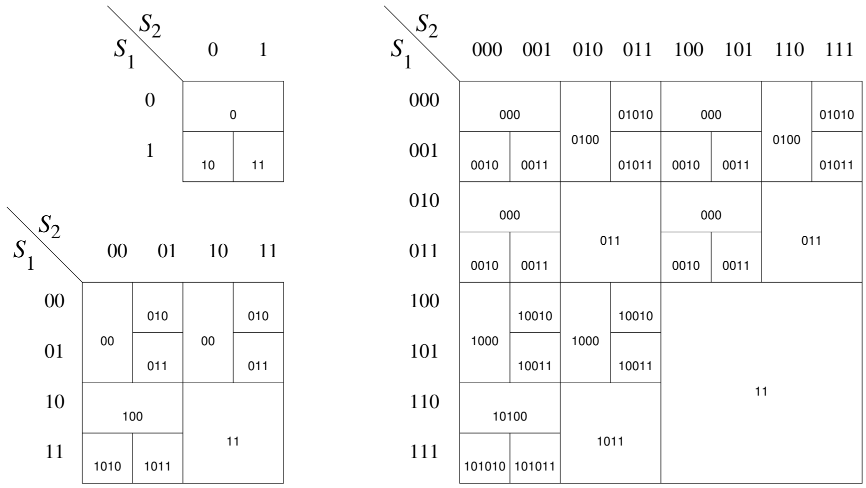

Figure 4: Partition of the value space for $k = 1$ (top left), 2 (bottom left), and 3 (right) induced by the alternating protocol for Disjointness_k; each rectangle is labeled with the transcript output by the protocol when run on inputs in the rectangle.

4.2.2 Objective PAR for specific protocols

Lemma 4.3. *If a protocol P for* Disjointness_k *tiles $f^{-1}(1)$ with 2^k tiles and tiles $f^{-1}(0)$ with $2^k - 1$ tiles, then the average-case objective PAR of P with respect to the uniform distribution equals*

$$2^k - 1 + \left(\frac{3}{4}\right)^k.$$

Proof. By the argument for Cor. 4.2, the contribution to this PAR value from those $S \subset f^{-1}(1)$ is $\left(\frac{3}{2}\right)^k$. The contribution to this PAR value from those $S \subset f^{-1}(0)$ is $4^{-k} \cdot (4^k - 3^k) \cdot (2^k - 1)$. Summing these together, we obtain the claimed value. □

Proposition 4.4. *The average-case objective PAR of the trivial protocol for* Disjointness_k *with respect to the uniform distribution is*

$$2^k - 1 + \left(\frac{3}{4}\right)^k.$$

Proof. The trivial protocol tiles $f^{-1}(1)$ with 2^k tiles (one for each set S_1 that player 1 might have), and it tiles $f^{-1}(0)$ with $2^k - 1$ tiles (one for each non-empty set S_1 that player 1 might have). We may then apply Lemma 4.3. □

Proposition 4.5. *The average-case objective PAR of the* 1*-first protocol for* Disjointness_k *with respect to the uniform distribution is*

$$2^k - 1 + \left(\frac{3}{4}\right)^k.$$

Proof. The protocol-induced tiles of $f^{-1}(1)$ correspond bijectively to the 2^k possible protocol transcripts $\{0, 10\}^k$, while the protocol-induced tiles of $f^{-1}(0)$ correspond bijectively to the 2^k-1 possible protocol transcripts $\{\{0, 10\}^i \times \{11\}\}_{i=0}^{k-1}$. We may then apply Lemma 4.3. □

Proposition 4.6. *The average-case objective PAR of the alternating protocol for* DISJOINTNESS$_k$ *with respect to the uniform distribution is*

$$2^k - 1 + \left(\frac{3}{4}\right)^k.$$

Proof. The protocol-induced tiles of $f^{-1}(1)$ correspond bijectively to the 2^k possible protocol transcripts $\{0, 10\}^k$, while the protocol-induced tiles of $f^{-1}(0)$ correspond bijectively to the 2^k-1 possible protocol transcripts $\{\{0, 10\}^i \times \{11\}\}_{i=0}^{k-1}$. We may then apply Lemma 4.3. □

4.3 Subjective PAR

4.3.1 Subjective PAR for the trivial protocol

Proposition 4.7. *The average-case PAR with respect to player* 1 *of the trivial protocol for* DISJOINTNESS$_k$ *is* 1. *The average-case PAR with respect to player* 2 *of the trivial protocol for* DISJOINTNESS$_k$*, and thus the average-case subjective PAR for the protocol, is*

$$2^k - 2\left(\frac{3}{2}\right)^k + 2\left(\frac{5}{4}\right)^k \sim 2^k \qquad (k \to \infty).$$

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

The 2-partition induced by the trivial protocol distinguishes between every pair of distinct inputs. To compute the average-case PAR with respect to player 2, we use v_k^0 and v_k^1 to denote the contributions (in the k-bit version of the problem) to the sum in Eq. 1 from tiles in $f^{-1}(0)$ and $f^{-1}(1)$, respectively, so the average-case PAR with respect to player 2 is then $\left(v_k^0 + v_k^1\right)/4^k$.

Let S be a 2-rectangle induced by the trivial protocol in the $k+1$-bit value space (so S is 1×1). If S is in either the bottom-left or the top-left quadrant, then the size of the ideal rectangle containing S is twice the size of the ideal rectangle that contains the corresponding induced rectangle in the k-bit value space (*i.e.*, the point in the k-bit space obtained by omitting the first bit of each input in S when the value space is depicted as in Fig. 1). This holds regardless of whether $S \subset f^{-1}(0)$ or $S \subset f^{-1}(1)$. If S is in the top-right quadrant and $S \subset f^{-1}(1)$, then the size of the ideal rectangle containing S is the same as that of the ideal rectangle containing the corresponding input in the k-bit value space; note that the bottom-right quadrant does not contain any points in $f^{-1}(1)$. If S is in the top-right quadrant and $S \subset f^{-1}(0)$, then the size of the ideal rectangle containing S is that of the ideal rectangle containing the corresponding input in the k-bit value space plus 2^k; the extra contribution of 2^k is added on for each of the $4^k - 3^k$ protocol-induced 2-rectangles in the top-right quadrant. If S is in the bottom-right quadrant (so that it is necessarily contained in $f^{-1}(0)$), then the size of the ideal rectangle containing S is at least 2^k (the part of the containing rectangle that is in the bottom-right quadrant); the amount by which this exceeds 2^k equals the size of the ideal 2-rectangle (for $f^{-1}(0)$) containing the corresponding point in the k-bit value space. In particular, each of the 2-rectangles for the k-bit value space is counted for exactly 2^k induced rectangles in the bottom-right quadrant, so the entire excess contribution is $2^k(4^k - 3^k)$.

We thus obtain the following recurrences (the terms are grouped by quadrant, clockwise from the bottom left).

$$\begin{array}{rcll} v^0_{k+1} & = & 2v^0_k + 2v^0_k + \left(v^0_k + 2^k(4^k - 3^k)\right) + \left(4^k \cdot 2^k + 2^k \cdot (4^k - 3^k)\right) & v^0_1 = 1 \\ v^1_{k+1} & = & 2v^1_k + 2v^1_k + v^1_k + 0 \qquad v^1_1 = 5 & \end{array}$$

From these, we obtain $v^1_k = 5^k$ and

$$v^0_k = 2^{3k} - 2^{1+k}3^k + 5^k,$$

from which it follows that the average-case subjective PAR with respect to player 2 (and thus for the trivial protocol) is

$$\frac{1}{4^k}(8^k - (2^{k+1}3^k) + 2 \cdot 5^k) = 2^k - 2\left(\frac{3}{2}\right)^k + 2\left(\frac{5}{4}\right)^k.$$

□

Corollary 4.8. *If* $\mathsf{PAR}^{\text{trivial}}_i$ *denotes the average-case PAR w.r.t. i of the trivial protocol for* DISJOINTNESS$_k$ *w.r.t. the uniform distribution, then*

$$\frac{\mathsf{PAR}^{\text{trivial}}_2}{\mathsf{PAR}^{\text{trivial}}_1} \sim 2^k \qquad (k \to \infty).$$

4.3.2 Subjective PAR for the 1-first protocol

Theorem 4.9. *The average-case PAR with respect to player 1 of the 1-first protocol for* DISJOINTNESS$_k$ *with respect to the uniform distribution is*

$$\frac{k}{2} - \frac{k}{3}\left(\frac{3}{4}\right)^k + \left(\frac{3}{4}\right)^k \sim \frac{k}{2} \qquad (k \to \infty).$$

The average-case PAR with respect to player 2 of the 1-first protocol for DISJOINTNESS$_k$ *with respect to the uniform distribution is*

$$\left(\frac{3}{2}\right)^k + \frac{1}{2}\left(\frac{5}{4}\right)^k - 1 + \frac{1}{2}\left(\frac{3}{4}\right)^k \sim \left(\frac{3}{2}\right)^k \qquad (k \to \infty).$$

Proof. To compute the average-case PAR with respect to player 1, we use h_k^0 and h_k^1 to denote the contributions (in the k-bit version of the problem) to the sum in Eq. 1 from the 1-induced tiles in $f^{-1}(0)$ and $f^{-1}(1)$, respectively, so the average-case PAR with respect to player 1 is then $\left(h_k^0 + h_k^1\right)/4^k$.

Let $S \subset f^{-1}(1)$ be a 1-rectangle induced by the 1-first protocol in the $(k+1)$-bit value space. If S is in the bottom-left quadrant, the ideal 1-rectangle containing S is the same size as the ideal rectangle that contains S in the k-bit value space (because there are no inputs in the bottom-right quadrant in $f^{-1}(1)$). If S is in one of the top quadrants, then the ideal 1-rectangle containing S is twice the size of the rectangle containing the rectangle that corresponds to S in the partition of the k-bit value space. Observe that each point in the top-left quadrant is in the same rectangle as the corresponding point in the top-right quadrant; in particular, this means that the induced 1-rectangles in the top two quadrants correspond bijectively to the induced 1-rectangles in the k-bit value space. S cannot be in the bottom-right quadrant, which contains no points in $f^{-1}(1)$. We thus have (separating the contributions of the bottom-left, top, and bottom-right quadrants)

$$h_{k+1}^1 = h_k^1 + 2h_k^1 + 0 = 3h_k^1.$$

By inspection, $h_1^1 = 1 + 2 + 0 = 3$; so $h_k^1 = 3^k$.

Now let $S \subset f^{-1}(0)$ be a 1-rectangle induced by the 1-first protocol in the $(k+1)$-bit value space. If S is in the bottom-left quadrant, then the size of the ideal 1-rectangle containing S equals the size of the ideal 1-rectangle containing S in the k-bit value space plus 2^k (because all of the inputs in the bottom-right quadrant in the same 1-rectangle as S are in the same ideal 1-rectangle as S). If nH_k^0 denotes the number of induced 1-rectangles $S \subset f^{-1}(0)$ in the bottom-left quadrant (this is the same as the total number of such 1-rectangles in the k-bit space), then the total extra contribution is $2^k nH_k^0$. If S is in the top two quadrants, the same arguments as before apply. If S is in the bottom-right quadrant (so that the size of S is 2^k), then the ideal 1-rectangle containing S has size 2^k plus the size of whatever part of the ideal 1-rectangle lies in the bottom-left quadrant. If we sum over all 2^k rectangles S in the bottom-right quadrant, the extra contribution from the bottom-left quadrant equals the total size of $f^{-1}(0)$ in the k-bit value space, *i.e.*, $4^k - 3^k$. This leads to (again separating the contributions of the bottom-left, top, and bottom-right quadrants)

$$h_{k+1}^0 = (h_k^0 + 2^k nH_k^0) + 2h_k^0 + ((4^k - 3^k) + 4^k).$$

By inspection, $h_1^0 = 1$. Because the bottom-left quadrant is a copy of the tiling of the k-bit space, the top two quadrants have the same number of rectangles, and the bottom-right quadrant has 2^k 1-rectangles, we have

$$nH_{k+1}^0 = nH_k^0 + nH_k^0 + 2^k,$$

with $nH_1^0 = 1$. From this, we obtain

$$nH_k^0 = k \cdot 2^{k-1},$$

which we then use to obtain

$$h_k^0 = \frac{k}{6}\left(3 \cdot 4^k - 2 \cdot 3^k\right).$$

Using PAR_1 to denote the average-case PAR with respect to 1, we have

$$\begin{aligned}
\mathsf{PAR}_1 &= \frac{1}{4^k}(h_k^0 + h_k^1) \\
&= \frac{1}{4^k}\left(\frac{k}{6}3 \cdot 4^k - \frac{k}{6}2 \cdot 3^k + 3^k\right) \\
&= \frac{k}{2} - \frac{k}{3}\left(\frac{3}{4}\right)^k + \left(\frac{3}{4}\right)^k
\end{aligned}$$

as claimed.

We now turn to the computation of the average-case PAR with respect to player 2. We use v_k^0 and v_k^1 to denote the contributions (in the k-bit version of the problem) to the sum in Eq. 1 from the 2-induced tiles in $f^{-1}(0)$ and $f^{-1}(1)$, respectively, so the average-case PAR with respect to player 2 is then $\left(v_k^0 + v_k^1\right)/4^k$.

Let $S \subset f^{-1}(1)$ be a 2-rectangle induced by the 1-first protocol in the $(k+1)$-bit value space. If S is in the bottom-left quadrant, the ideal 2-rectangle containing S is twice as big as the ideal 2-rectangle that contains S in the k-bit value space. The same holds true if S is in the top-left quadrant. If S is in the top-right quadrant, the ideal 2-rectangle containing S is the same size as in the k-bit value space. Finally, the bottom-right quadrant does not contain any values in $f^{-1}(1)$. Thus, we have (again listing contributions clockwise from the bottom-left quadrant)

$$v_{k+1}^1 = 2v_k^1 + 2v_k^1 + v_k^1 + 0 = 5v_k^1.$$

By inspection, $v_1^1 = 5$; so, $v_k^1 = 5^k$.

Now let $S \subset f^{-1}(0)$ be a 2-rectangle induced by the 1-first protocol in the $(k+1)$-bit value space. If S is in the bottom-left or top-left quadrant, the ideal 2-rectangle containing S is twice as big as in the k-bit value space. If S is in the top-right quadrant, the size of the ideal 2-rectangle containing S equals 2^k plus the size of the ideal 2-rectangle that contains S in the k-bit value space. Finally, if we sum over all S in the bottom-right quadrant, the total sizes of the ideal 2-rectangles containing these S is $2^k \cdot 2^k$ plus the total size of $f^{-1}(0)$ in the k-bit value space. Combining all of these relations, and using nV_k^0 to denote the number of 2-rectangles in $f^{-1}(0)$ in the k-bit value space, we have

$$v_{k+1}^0 = 2v_k^0 + 2v_k^0 + (v_k^0 + 2^k \cdot nV_k^0) + (4^k + 4^k - 3^k).$$

(As above, contributions are grouped by quadrant clockwise from the bottom right.) We also have

$$nV_{k+1}^0 = nV_k^0 + nV_k^0 + nV_k^0 + 2^k = 3nV_k^0 + 2^k.$$

By inspection, $v_1^0 = 1$ and $nV_1^0 = 1$. From this, we obtain

$$nV_k^0 = 3^k - 2^k$$

and then

$$v_k^1 = -4^k + \frac{1}{2}3^k + 6^k - \frac{1}{2}5^k.$$

Using PAR_2 to denote the average-case PAR with respect to 2, we have

$$\begin{aligned}
\mathsf{PAR}_2 &= \frac{1}{4^k}(v_k^0 + v_k^1) \\
&= \frac{1}{4^k}\left(5^k + 6^k - \frac{1}{2}5^k - 4^k + \frac{1}{2}3^k\right) \\
&= \left(\frac{3}{2}\right)^k + \frac{1}{2}\left(\frac{5}{4}\right)^k - 1 + \frac{1}{2}\left(\frac{3}{4}\right)^k
\end{aligned}$$

as claimed. □

Corollary 4.10. *The average-case subjective PAR of the* 1*-first protocol for* DISJOINTNESS$_k$ *with respect to the uniform distribution is*

$$\left(\frac{3}{2}\right)^k + \frac{1}{2}\left(\frac{5}{4}\right)^k - 1 + \frac{1}{2}\left(\frac{3}{4}\right)^k \sim \left(\frac{3}{2}\right)^k \qquad (k \to \infty).$$

Corollary 4.11. *If* $\mathsf{PAR}_i^{1-\text{first}}$ *denotes the average-case PAR w.r.t.* i *of the* 1*-first protocol for* DISJOINTNESS$_k$ *w.r.t. the uniform distribution, then*

$$\frac{\mathsf{PAR}_2^{1-\text{first}}}{\mathsf{PAR}_1^{1-\text{first}}} \sim \frac{2}{k}\left(\frac{3}{2}\right)^k \qquad (k \to \infty).$$

4.3.3 Subjective PAR for the alternating protocol

We let PAR_k^i denote the PAR w.r.t. i for the alternating protocol for DISJOINTNESS$_k$. We let h_k^1 and v_k^1 be the contributions of $f^{-1}(1)$ to the sums analogous to that in Eq. 1 for objective PAR, *i.e.*,

$$h_k^1 = \sum_{S \subseteq f^{-1}(1)} |R^I(S)| \qquad v_k^1 = \sum_{T \subseteq f^{-1}(1)} |R^I(T)|,$$

where the sum for h_k^1 is taken over protocol-induced "horizontal" rectangles S (in the induced 1-partition) on which f takes the value 1, and the sum for v_k^1 is taken over protocol-induced "vertical" rectangles T (in the induced 2-partition) on which f takes the value 1. Using the structure of the induced tiling, we may obtain recurrences for h_k^1 and v_k^1 as follows.

$$h_k^1 = h_{k-1}^1 + 2v_{k-1}^1 + 0 \qquad h_1^1 = 3 \tag{2}$$

$$v_k^1 = 2v_{k-1}^1 + \left(2h_{k-1}^1 + h_{k-1}^1\right) + 0 \qquad v_1^1 = 5 \tag{3}$$

In each recurrence, the first summand is the contribution from the bottom-left quadrant, the second summand is the contribution from the two top quadrants, and the third summand is the contribution from the bottom-right quadrant. From these recurrences, we obtain $h_k^1 = \frac{4}{5}2^{2k} + \frac{(-1)^k}{5}$ and $v_k^1 = \frac{6}{5}2^{2k} - \frac{(-1)^k}{5}$.

We define h_k^0 and v_k^0 analogously to capture the contributions of $f^{-1}(0)$ to the sums under consideration; we will also keep track of the number of tiles in the 1- and 2-induced partitions on which f takes the value 0 (the "horizontal" and "vertical" tiles, which we denote as nH_k^0 and nV_k^0, respectively).

We start with the following recurrences for nH_k^0 and nV_k^0.

$$\begin{aligned} nH_{k+1}^0 &= nH_k^0 + nV_k^0 + 2^k & nH_1^0 = 1 \\ nV_{k+1}^0 &= nV_k^0 + \left(nH_k^0 + nH_k^0\right) + 2^k & nV_1^0 = 1 \end{aligned}$$

From these, we obtain

$$\begin{aligned} nH_k^0 &= -2^{k+1} + (1 - 3/(2\sqrt{2})) \cdot (1-\sqrt{2})^k + ((1+\sqrt{2})^k \cdot (4+3\sqrt{2}))/4 \\ nV_k^0 &= -3 \cdot 2^k + (1-\sqrt{2})^k \cdot (3/2 - \sqrt{2}) + (1+\sqrt{2})^k \cdot (3/2+\sqrt{2}) \end{aligned}$$

We obtain the following recurrences for h_k^0 and v_k^0.

$$\begin{aligned} h_{k+1}^0 &= \left(h_k^0 + nH_k^0 \cdot 2^k\right) + 2v_k^0 + \left(4^k + 4^k - 3^k\right) & h_1^0 = 1 \\ v_{k+1}^0 &= 2v_k^0 + 2h_k^0 + \left(h_k^0 + nH_k^0 \cdot 2^k\right) + \left(4^k + 4^k - 3^k\right) & v_1^0 = 1 \end{aligned}$$

From these, we may obtain

$$h_k^0 = \frac{1}{20\sqrt{2}}\Bigg(5 \cdot 2^{k+1} \cdot (1-\sqrt{2})^k \cdot (-3+2\sqrt{2}) + 5 \cdot 2^{k+1} \cdot (1+\sqrt{2})^k \cdot (3+2\sqrt{2}) \\ + \sqrt{2}((-1)^k - 7 \cdot 2^{2k+3} + 5 \cdot 3^{k+1})\Bigg)$$

$$v_k^0 = \frac{1}{20}\left(-(-1)^k + 25 \cdot 3^k - 21 \cdot 4^{k+1} - 5 \cdot 2^{k+1}(1-\sqrt{2})^k(-3+2\sqrt{2}) + 5 \cdot 2^{k+1}(1+\sqrt{2})^k(3+2\sqrt{2})\right)$$

We may now compute the PAR with respect to each of the two players as

$$\mathsf{PAR}_1^{\mathrm{alt}}(k) = \frac{h_k^0 + h_k^1}{2^{2k}} \qquad \text{and} \qquad \mathsf{PAR}_2^{\mathrm{alt}}(k) = \frac{v_k^0 + v_k^1}{2^{2k}}.$$

Theorem 4.12. *The average-case PAR with respect to player* 1 *of the alternating protocol for* DISJOINTNESS$_k$ *with respect to the uniform distribution is*

$$\mathsf{PAR}_1^{\mathrm{alt}}(k) = \frac{1}{4^{k+1}}\Bigg((-1)^k - 2^{2k+3} + 3^{k+1} \\ + (4-3\sqrt{2})(2-2\sqrt{2})^k + (2+2\sqrt{2})^k(4+3\sqrt{2})\Bigg) \\ \sim \frac{4+3\sqrt{2}}{4}\left(\frac{1+\sqrt{2}}{2}\right)^k \qquad (k \to \infty)$$

The average-case PAR with respect to player 2 *of the alternating protocol for* DISJOINTNESS$_k$ *with respect to the uniform distribution is*

$$\begin{aligned}\mathsf{PAR}_2^{\text{alt}}(k) = \frac{1}{4^{k+1}}\Big(&-(-1)^k + 5\cdot 3^k - 3\cdot 4^{k+1} \\ &+ 2^{k+1}(3-2\sqrt{2})(1-\sqrt{2})^k + 2^{k+1}(3+2\sqrt{2})(1+\sqrt{2})^k\Big) \\ &\sim \frac{3+2\sqrt{2}}{2}\left(\frac{1+\sqrt{2}}{2}\right)^k \qquad (k\to\infty)\end{aligned}$$

Corollary 4.13. *The average-case subjective PAR of the alternating protocol for* DISJOINTNESS$_k$ *with respect to the uniform distribution is*

$$\begin{aligned}&\frac{1}{4^{k+1}}\Big(-(-1)^k + 5\cdot 3^k - 3\cdot 4^{k+1} + 2^{k+1}(3-2\sqrt{2})(1-\sqrt{2})^k + 2^{k+1}(3+2\sqrt{2})(1+\sqrt{2})^k\Big) \\ &\qquad\sim \frac{3+2\sqrt{2}}{2}\left(\frac{1+\sqrt{2}}{2}\right)^k \qquad (k\to\infty)\end{aligned}$$

Corollary 4.14. *If* $\mathsf{PAR}_i^{\text{alt}}(k)$ *denotes the average-case PAR w.r.t.* i *of the* 1*-first protocol for* DISJOINTNESS$_k$ *w.r.t. the uniform distribution, then*

$$\frac{\mathsf{PAR}_2^{\text{alt}}(k)}{\mathsf{PAR}_1^{\text{alt}}(k)} \sim \sqrt{2} \qquad (k\to\infty).$$

5 PARs for INTERSECTION$_k$

5.1 Structure of Protocol-Induced Tilings

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

Lemma 5.1. *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 5 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. 5, the bottom-left quadrant is

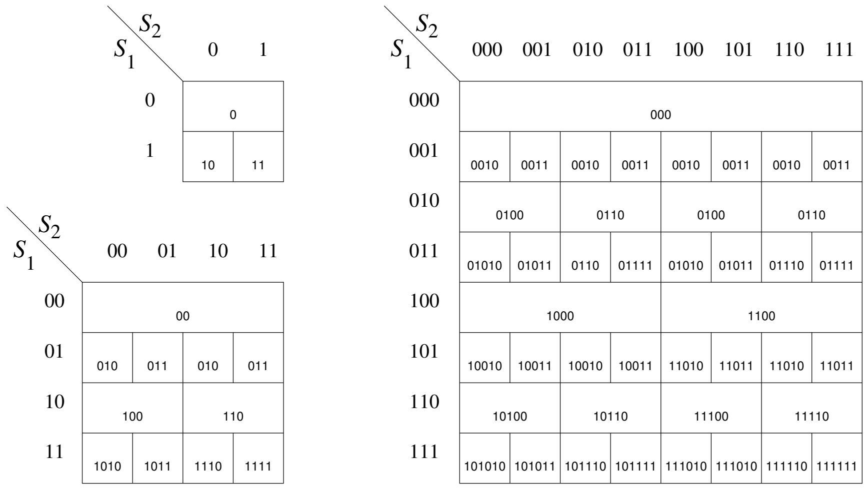

Figure 5: Partition of the value space for $k = 1$ (top left), 2 (bottom left), and 3 (right) induced by the trivial and 1-first protocols for $\textsc{Intersection}_k$; each rectangle is labeled with the transcript output by the protocol when run on inputs in the rectangle.

$10T_k$ (*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 6 depicts the tilings of the 1-, 2-, and 3-bit value spaces induced by the alternating protocol for $\textsc{Intersection}_k$. 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. 6, 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 $11T_k$.

5.2 Objective PAR

5.2.1 Lower bound

We obtain the following result for the average-case objective PAR of the $\textsc{Intersection}_k$ problem.

Theorem 5.2. *The average-case objective PAR of the* $\textsc{Intersection}_k$ *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

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

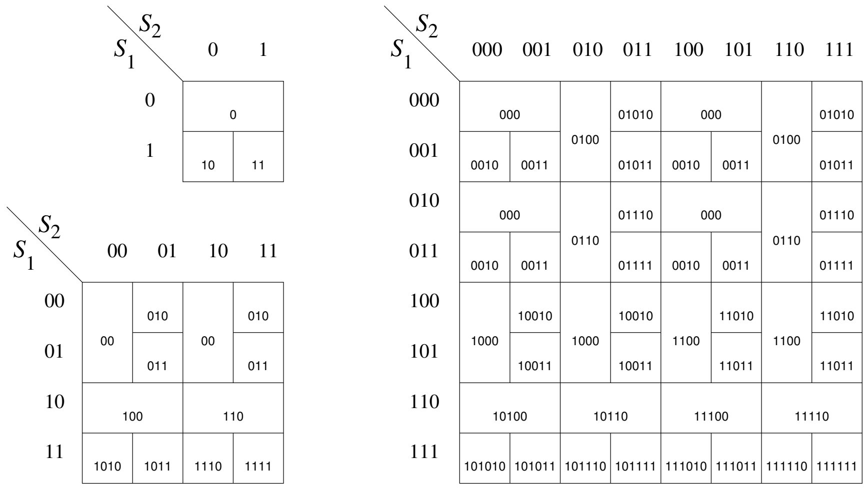

Figure 6: Partition of the value space for $k = 1$ (top left), 2 (bottom left), and 3 (right) induced by the alternating protocol for $\textsc{Intersection}_k$; each rectangle is labeled with the transcript output by the protocol when run on inputs in the rectangle.

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. 2) 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. 2). 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 $k = 1$, shown in Fig. 7, 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

$$\mathsf{PAR}_1 = \frac{1}{2^2}(3+3+1) = \frac{7}{4}.$$

□

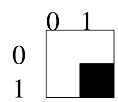

Figure 7: Ideal partition for the $\textsc{Intersection}_k$ problem with $k = 1$.

5.2.2 Objective PAR for the trivial and 1-first protocols

Proposition 5.3. *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. □

5.2.3 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. 5.3 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 5.4. *The average-case objective PAR for the alternating protocol for the* Intersection$_k$ *problem equals* $\left(\frac{7}{4}\right)^k$. □

5.3 Subjective PAR

5.3.1 Subjective PAR for the trivial and 1-first protocols

Remark 5.5. The contribution from $f^{-1}(\emptyset)$ is as for Disjointness$_k$. What about the contribution for $f^{-1}(\neq \emptyset)$?

Proposition 5.6. *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 5.7. *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 5.8. *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*

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

5.3.2 Subjective PAR for the alternating protocol

Proposition 5.9. *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

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

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 $v_k = 6 \cdot 5^{k-1}$. Denoting by $\mathsf{PAR}_i^{\text{alt}}(k)$ the average-case PAR w.r.t. i of the trivial protocol for INTERSECTION$_k$ w.r.t. the uniform distribution, we have

$$\begin{aligned} \mathsf{PAR}_1^{\text{alt}}(k) &= \frac{h_k}{4^k} = \frac{4}{5}\left(\frac{5}{4}\right)^k \\ \mathsf{PAR}_2^{\text{alt}}(k) &= \frac{v_k}{4^k} = \frac{6}{5}\left(\frac{5}{4}\right)^k \end{aligned}$$

as claimed. □

Corollary 5.10. *The average-case subjective PAR of the alternating protocol for* INTERSECTION$_k$ *is*

$$\frac{6}{5}\left(\frac{5}{4}\right)^k.$$

Corollary 5.11. *If* $\mathsf{PAR}_i^{\text{alt}}$ *denotes the average-case PAR w.r.t.* i *of the trivial protocol for* INTERSECTION$_k$ *w.r.t. the uniform distribution, then*

$$\frac{\mathsf{PAR}_2^{\text{alt}}}{\mathsf{PAR}_1^{\text{alt}}} = \frac{3}{2}.$$

6 Conclusions and Future Work

Our definitions of PARs involve 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. For example, if there is a natural notion of "distance" between inputs (as in the examples considered in this paper), one might prefer protocols that cannot distinguish among a few inputs that are far from each other to protocols that cannot distinguish among many inputs that are all relatively close. This necessitates different definitions of PARs and suggests many interesting avenues for future work.

Starting from the same place that we did, namely [4, 11], Bar-Yehuda *et al.* [1] provided three definitions of approximate privacy. We show in [8] that the formulation in [1] is not equivalent to ours, but there is more to do along these lines. The definition in [1] that seems most relevant to the study of privacy-approximation ratios is their notion of *h-privacy*. Determine when and how it is possible to express PARs in terms of h-privacy and *vice versa*.

Lower bounds on the average-case subjective PARs for $\textsc{Disjointness}_k$ and $\textsc{Intersection}_k$ would be interesting; as noted above, we conjecture that these are exponential in k. Our PAR framework should also be applied to other functions and extended to n-party communication.

References

[1] Reuven Bar-Yehuda, Benny Chor, Eyal Kushilevitz, and Alon Orlitsky. Privacy, additional information, and communication. *IEEE Transactions on Information Theory*, 39:55–65, 1993.

[2] Amos Beimel, Paz Carmi, Kobbi Nissim, and Enav Weinreb. Private approximation of search problems. In *Proceedings of the ACM Symposium on Theory of Computing*, pages 119–128, 2006.

[3] Felix Brandt and Tuomas Sandholm. On the existence of unconditionally privacy-preserving auction protocols. *ACM Trans. Inf. Syst. Secur.*, 11(2):1–21, 2008.

[4] Benny Chor and Eyal Kushilevitz. A zero-one law for boolean privacy. *SIAM J. Discrete Math*, 4:36–47, 1991.

[5] Yevgeniy Dodis, Shai Halevi, and Tal Rabin. A cryptographic solution to a game-theoretic problem. In *Advances in Cryptology—Crypto 2000*, Springer Verlag LNCS.

[6] Cynthia Dwork. Differential privacy. In *Proceedings of the International Colloquium on Automata, Languages and Programming*, pages 1–12, 2006.

[7] Joan Feigenbaum, Yuval Ishai, Tal Malkin, Kobbi Nissim, Martin J. Strauss, and Rebecca N. Wright. Secure multiparty computation of approximations. *ACM Transactions on Algorithms*, 2(3):435–472, 2006.

[8] Joan Feigenbaum, Aaron D. Jaggard, and Michael Schapira. Approximate Privacy: Foundations and Quantification. Technical Report 2009-14, DIMACS, 2009. `http://dimacs.rutgers.edu/TechnicalReports/abstracts/2009/2009-14.html` and arXiv:0910.5714.

[9] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proceedings of the ACM Symposium on Theory of Computing*, pages 351–360, 2009.

[10] Shai Halevi, Robert Krauthgamer, Eyal Kushilevitz, and Kobbi Nissim. Private approximation of NP-hard functions. In *Proceedings of the ACM Symposium on Theory of Computing*, pages 550–559, 2001.

[11] Eyal Kushilevitz. Privacy and communication complexity. *SIAM J. Discrete Math.*, 5(2):273–284, 1992.

[12] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[13] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, pages 94–103, 2007.

[14] Moni Naor, Benny Pinkas, and Reuben Sumner. Privacy preserving auctions and mechanism design. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 129–139, 1999.

[15] Andrew C. Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the ACM Symposium on Theory of Computing*, pages 209–213, 1979.

A Perfect Privacy and Communication Complexity

For convenience, we include Sec. 2 of (a revised version of) [8] as the text of this appendix. It contains the basic definitions of communication complexity and privacy that underlie our approach to approximate privacy.

A.1 Two-Party Communication Model

We now briefly review Yao's model of two-party communication and notions of objective and subjective perfect privacy; see Kushilevitz and Nisan [12] 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.

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 $f : \{0,1\}^k \times \{0,1\}^k \rightarrow \{0,1\}^t$. 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 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 [11], 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 $f : \{0,1\}^k \times \{0,1\}^k \rightarrow \{0,1\}^t$ 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 A.1 (Regions, partitions)**.** A *region* in a matrix A is any subset of entries in A (not necessarily a submatrix). A *partition* of A is a collection of disjoint regions in A whose union equals A.

Definition A.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 A.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 A.4 (Refinements)**.** A tiling $T_1(f)$ of a matrix $A(f)$ is said to be a *refinement* of another tiling $T_2(f)$ of $A(f)$ if every rectangle in $T_1(f)$ is contained in some rectangle in $T_2(f)$.

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 $i \in \{0, 1\}$ 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*".

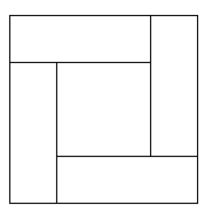

Figure 8: A tiling that is not induced by any communication protocol [11]

Remark A.5. There are monochromatic tilings that cannot be induced by communication protocols. For example, observe that the tiling in Fig. 8 (which is essentially an example from [11]) has this property.

A.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. (This 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 (that 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 A.6 (Perfect privacy with respect to 1)**.** [4,11] P is *perfectly private with respect to party* 1 if, for every 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. A.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 A.7 (Perfect privacy with respect to 2)**.** [4,11] P is *perfectly private with respect to party* 2 if, for every x_1, x_1' 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 A.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 A.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 A.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 $s_{(x_1,x_2)} = s_{(x_1',x_2')}$.

Kushilevitz [11] 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 A.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 A.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 A.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.