On Communication Protocols that Compute Almost Privately

Marco Comi[1], Bhaskar DasGupta[1], Michael Schapira[2], and Venkatakumar Srinivasan[1]

[1] Department of Computer Science, University of Illinois at Chicago, IL 60607
ingmarco85@gmail.com, {bdasgup,vsrini7}@uic.edu

[2] Department of Computer Science, Princeton University, Princeton, NJ 08540
ms7@cs.princeton.edu

Abstract. A traditionally desired goal when designing auction mechanisms is *incentive compatibility*, *i.e.*, ensuring that bidders fare best by truthfully reporting their preferences. A *complementary* goal, which has, thus far, received significantly less attention, is to *preserve privacy*, *i.e.*, to ensure that bidders reveal no more information than necessary. We further investigate and generalize the approximate privacy model for two-party communication recently introduced by Feigenbaum *et al.* [8]. We explore the privacy properties of a natural class of communication protocols that we refer to as "*dissection protocols*". Dissection protocols include, among others, the bisection auction in [9, 10] and the bisection protocol for the millionaires problem in [8]. Informally, in a dissection protocol the communicating parties are restricted to answering simple questions of the form "*Is your input between the values* α *and* β (under a pre-defined order over the possible inputs)?".

We prove that for a large class of functions called *tiling functions*, which include the 2$^{\text{nd}}$-price Vickrey auction, there *always* exists a dissection protocol that provides a *constant average-case privacy approximation ratio* for uniform or "almost uniform" probability distributions over inputs. To establish this result we present an interesting connection between the approximate privacy framework and basic concepts in computational geometry. We show that such a good privacy approximation ratio for tiling functions does *not*, in general, exist in the *worst case*. We also discuss extensions of the basic setup to more than two parties and to non-tiling functions, and provide calculations of privacy approximation ratios for two functions of interest.

Keywords: Approximate Privacy, Auctions, Communication Protocols

1 Introduction

Consider the following interaction between two parties, Alice and Bob. Each of the two parties, Alice and Bob, holds a *private* input, x_{bob} and y_{alice} respectively, not known to the other party. The two parties aim to compute a function f of the two private inputs. Alice and Bob alternately query each other to make available a *small* amount of information about their private inputs, *e.g.*, an answer to a range query on their private inputs or a few bits of their private inputs. This process ends when each of them has seen enough information to be able to compute the value of $f(x_{\text{bob}}, y_{\text{alice}})$. The central question that is the focus of this paper is:

> *Can we design a communication protocol whose execution reveals, to both Alice and Bob, as well as to any eavesdropper, as little information as possible about other the other's private input beyond what is necessary to compute the function value*?

Note that there are two conflicting constraints: Alice and Bob need to communicate sufficient information for computing the function value, but would prefer not to communicate too much information about their private inputs. This setting can be generalized in an obvious manner to $d > 1$ parties $\mathsf{party}_1, \mathsf{party}_2, \ldots, \mathsf{party}_d$ computing a d-ary f by querying the parties in round-robin order, allowing each party to broadcast information about its private input (via a public communication channel).

Privacy preserving computational models such as the one described above have become an important research area due to the increasingly widespread usage of sensitive data in networked environments, as evidenced by distributed computing applications, game-theoretic settings (*e.g.*, auctions) and more. Over the years computer scientists have explored many *quantifications* of privacy in computation. Much of this research focused on designing *perfectly* privacy-preserving protocols, *i.e.*, protocols whose execution reveals *no* information about the parties' private inputs beyond that implied by the outcome of the computation. Unfortunately, perfect privacy is often either *impossible*, or *infeasibly costly* to achieve. To overcome this, researchers have also investigated various notions of *approximate privacy* [7, 8].

In this paper, we adopt the approximate privacy framework of [8] that quantifies approximate privacy via the *privacy approximation ratios* (Pars) of protocols for computing a deterministic function of two private inputs. *Informally*, Par captures the objective that an observer of the transcript of the entire protocol will not be able to distinguish the real inputs of the two communicating parties from *as large a set as possible* of other inputs. To capture this intuition, [8] makes use of the machinery of communication-complexity theory to provide a geometric and combinatorial interpretation of protocols. [8] formulates both the worst-case and the average-case version of Pars and studies the tradeoff between privacy preservation and communication complexity for several functions.

1.1 Economic Motivation

The original motivation of this line of research, as explained in [8], comes from privacy concerns in auction theory. A traditionally desired goal when designing an auction mechanism is to ensure that it is *incentive compatible*, *i.e.*, bidders fare best by truthfully reporting their preferences. More recently, attention has also been given to the *complementary* goal of preserving the privacy of the bidders (both with respect to each other and to the auctioneer/mechanism). Take, for example, the famous 2^{nd}-price Vickrey auction of an item. Consider the ascending-price English auction, *i.e.*, the straightforward protocol in which the price of the item is incrementally increased, bidders drop out when their value for the item is exceeded until the identity of winner is determined, and the winner is then charged the second-highest bid. Intuitively, this protocol reveals more information than what is *absolutely necessary* to compute the outcome, *i.e.*, the identity of the winner and the second-highest bid. Specifically, observe under the ascending-price English auction not only will the value of the second-highest bidder be revealed, but so will the values of all other bidders but the winner.

Can we design communication protocols which implement the 2^{nd}-price Vickrey auction in an (approximately) privacy-preserving manner? Can we design such protocols that are computationally- or communication-efficient? These sort of questions motivate our work. We consider a setting that captures applications of the above type, and explore the privacy-preservation and communication-complexity guarantees achievable in this setting.

2 Summary of Our Contributions

Any investigation of approximate privacy for multi-party computation starts by defining how we quantify approximate privacy. In this paper, we use the combinatorial framework of [8] for quantification of approximate privacy for two parties via PARs and present its natural extension to three or more parties. Often, parties' inputs have a natural ordering, *e.g.*, the private input of a party belongs to some range of integers $\{L, L+1, \ldots, M\}$ (as is the case when computing, say, the maximum or minimum of two inputs). When designing protocols for such environments, a natural restriction is to only allow the protocol to ask each party questions of the form "*Is your input between the values α and β (under this natural order over possible inputs)?*". We refer to this type of protocols as *dissection protocols* and study the privacy properties of this natural class of protocols. We note that the bisection and c-bisection protocols for the millionaires problem and other problems in [8], as well as the bisection auction in [9, 10], all fall within this category of protocols. Our findings are summarized below.

Average- and worst-case PARs for tiling functions for two party computation. We first consider a broad class of functions, namely the *tiling functions*, that encompasses several well-studied functions (*e.g.*, Vickrey's second-price auctions). Informally, a two-variable tiling function is a function whose output space can be viewed as a collection of disjoint combinatorial rectangles in the two-dimensional plane, where the function has the same value within each rectangle. A first natural question for investigation is to classify those tiling functions for which there exists a perfectly privacy-preserving dissection protocol. We observe that for every Boolean tiling functions (*i.e.*, tiling functions which output binary values) *this is indeed the case.* In contrast, for tiling functions with a range of just three values, perfectly privacy-preserving computation is no longer necessarily possible (even when not restricted to dissection protocols).

We next turn our attention to PARs. We prove that for *every* tiling function there exists a dissection protocol that achieves a constant PAR in the average case (that is, when the parties' private values are drawn from an uniform or *almost* uniform probability distribution). To establish this result, we make use of results on the binary space partitioning problems studied in the computational geometry literature. We complement this positive result for dissection protocols with the following negative result: *there exist tiling functions for which no dissection protocol can achieve a constant* PAR *in the worst-case.*

Extensions to non-tiling functions and three-party communication. We discuss two extensions of the above results. We explain how our constant average-case PAR result for tiling functions can be extended to a family of "almost" tiling functions. In addition, we consider the case of *more than two* parties. We show that in this setting it is *no longer true* that for every tiling function there exists a dissection protocol that achieves a constant PAR in the average case. Namely, we exhibit a three-dimensional tiling function for which *every* dissection protocol exhibits *exponential* average- and worst-case PARs, *even when an unlimited number of communication steps is allowed.*

PARs for the set covering and equality functions. [8] presents bounds on the average-case and the worst-case PARs of the bisection protocol — a special case of dissection protocols — for several functions. We analyze the PARs of the bisection protocol for two well-studied Boolean functions: the set-covering and equality functions; the equality function provides a useful testbed for evaluating privacy preserving protocols [3] [11, Example 1.21] and set-covering type of functions are useful for studying the differences between deterministic and non-deterministic communication complexities [11]. We show that, for both functions, the bisection protocol *fails to achieve* good PARs in both the average- and the worst-case.

3 Summary of Prior Related Works

3.1 Privacy-preserving Computation

Privacy-preserving computation has been the subject of extensive research and has been approached from information-theoretic [3], cryptographic [5], statistical [12], communication complexity [13, 16], statistical database query [7] and other perspectives [11]. Among these, most relevant to our work is the approximate privacy framework of Feigenbaum *et al.* [8] that presents a metric for quantifying privacy preservation building on the work of Chor and Kushilevitz [6] on characterizing perfectly privately computable computation and on the work of Kushilevitz [13] on the communication complexity of perfectly private computation. The bisection, c-bisection and bounded bisection protocols of [8] fall within our category of dissection protocol since we allow the input space of each party to be divided into two subsets of arbitrary size. There are also some other formulations of perfectly and approximately privacy-preserving computation in the literature, but they are inapplicable in our context. For example, the differential privacy model (see [7]) approaches privacy in a different context via adding noise to the result of a database query in such a way as to preserve the privacy of the individual records but still have the result convey nontrivial information,

3.2 Binary space partition (Bsp)

Bsps present a way to implement a *geometric divide-and-conquer* strategy and is an extremely popular approach in numerous applications such as hidden surface removal, ray-tracing, visibility problems, solid geometry, motion planning and spatial databases. However, to the best of our knowledge, a connection between Bsps bounds such as in [2, 4, 14, 15] and approximate privacy has not been explored before.

4 The Model and Basic Definitions

4.1 Two-party Approximate Privacy Model of [8]

We have two parties party_1 and party_2, having binary strings x_1 and x_2 respectively, which represents their private values in some set $\mathcal{U}^{\text{in}}$. The common goal of the two parties is to compute the value $f(x_1, x_2)$ of a given public-knowledge function f. Before a communication protocol P starts, each party_i initializes its "*set of maintained inputs*" $\mathcal{U}_i^{\text{in}}$ to $\mathcal{U}^{\text{in}}$. In one step of communication, one party transmits a bit indicating in which of two parts of its input space its private input lies. The other party then updates its set of maintained inputs accordingly. The very last information transmitted in the protocol P contains the value of of $f(x_1, x_2)$. The final transcript of the protocol is denoted by $s(x_1, x_2)$.

Denoting the domain of outputs by $\mathcal{U}^{\text{out}}$, any function $f : \mathcal{U}^{\text{in}} \times \mathcal{U}^{\text{in}} \mapsto \mathcal{U}^{\text{out}}$ can be visualized as $|\mathcal{U}^{\text{in}}| \times |\mathcal{U}^{\text{in}}|$ matrix with entries from $\mathcal{U}^{\text{out}}$ in which the first dimension represents the possible values of party_1, ordered by some permutation Π_1, while the second dimension represents the possible values of party_2, ordered by some permutation Π_2, and each entry contains the value of f associated with a particular set of inputs from the two parties. This matrix will be denoted by $A_{\Pi_1,\Pi_2}(f)$, or sometimes simply by A. We present the following definitions from [8, 11].

Definition 1 (Regions, partitions) *A region of A is any subset of entries in A. A partition of A is a collection of disjoint regions in A whose union is A.*

Definition 2 (Rectangles, tilings, refinements) *A rectangle in A is a submatrix of A. A tiling of A is a partition of A into rectangles. A tiling T_1 of A is a refinement of another tiling T_2 of A if every rectangle in T_1 is contained in some rectangle in T_2.*

Definition 3 (Monochromatic, maximal monochromatic and ideal monochromatic partitions) *A region R of A is monochromatic if all entries in R are of the same value. A monochromatic partition of A is a partition with only monochromatic regions. A monochromatic region of A is a maximal monochromatic region if no monochromatic region in A properly contains it. The ideal monochromatic partition of A consists of the maximal monochromatic regions.*

Definition 4 (Perfect privacy) *Protocol P achieves perfect privacy if, for every two sets 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 $s(x_1,\ x_2) = s(x'_1,\ x'_2)$. Equivalently, a protocol P for f achieves perfectly privacy if the monochromatic tiling induced by P is the ideal monochromatic partition of $A(f)$.*

Definition 5 (Worst case and average case Par of a protocol P) *Let $R^P(x_1, x_2)$ be the monochromatic rectangle containing the cell $A(x_1, x_2)$ induced by P, $R^I(x_1, x_2)$ be the monochromatic region containing the cell $A(x_1, y_1)$ in the ideal monochromatic partition of A, and $\mathcal{D}$ be a probability distribution over the space of inputs. Then P has a worst-case* Par *of α_{worst} and an average case* Par *of $\alpha_{\mathcal{D}}$ under distribution $\mathcal{D}$ provided*[3]

$$\alpha_{\text{worst}} = \max_{(x_1,x_2)\in\mathcal{U}^{\text{in}}\times\mathcal{U}^{\text{in}}} \frac{|\, R^I(x_1,x_2)\,|}{|\, R^P(x_1,x_2)|} \quad \textit{and} \quad \alpha_{\mathcal{D}} = \sum_{(x_1,x_2)\in\mathcal{U}^{\text{in}}\times\mathcal{U}^{\text{in}}} \Pr_{\mathcal{D}}[x_1 \,\&\, x_2]\, \frac{\left|R^I(x_1,x_2)\right|}{|R^P(x_1,x_2)|}$$

Definition 6 (Par for a function) *The worst-case (average-case)* Par *for a function f is the minimum, over all protocols P for f, of the worst-case (average-case)* Par *of P.*

Extension to Multi-party Computation In the multi-party setup, we have $d > 2$ parties $\mathsf{party}_1, \mathsf{party}_2, \ldots, \mathsf{party}_d$ computing a d-ary function $f : (\mathcal{U}^{\text{in}})^d \mapsto \mathcal{U}^{\text{out}}$. Now, f can be visualized as $\left|\mathcal{U}^{\text{in}}\right| \times \cdots \times \left|\mathcal{U}^{\text{in}}\right|$ matrix $A_{\Pi_1,\ldots,\Pi_d}(f)$ (or, sometimes simply by A) with entries from $\mathcal{U}^{\text{out}}$ in which the i^{th} dimension represents the possible values of party_i ordered by some permutation Π_i, and each entry of A contains the value of f associated with a particular set of inputs from the d parties. Then, all the previous definitions can be naturally adjusted in the obvious manner, *i.e.*, the input space as a d-dimensional space, each party maintains the input partitions of all other $d-1$ parties, the transcript of the protocol s is a d-ary function, and rectangles are replaced by d-dimensional hyper-rectangles (Cartesian product of d intervals).

4.2 Dissection Protocols & Tiling Functions for 2-party Computation

Often in a communication complexity settings the input of each party has a natural ordering, *e.g.*, the set of input of a party from $\{0,1\}^k$ can represent the numbers $0, 1, 2, \ldots, 2^k - 1$ (as is the case when computing the maximum/minimum of two inputs, in the millionaires problem, in second-price auctions, and more). When designing protocols for such environments, a natural restriction is to only the allow protocols such that each party asks questions of the form "*Is your input between a and b (in this natural order over possible inputs)*?", where $a, b \in \{0,1\}^k$. Notice that after applying an appropriate permutation to the inputs, such a protocol divides the input space into two (not necessarily equal) halves. Below, we formalize these types of protocols as "*dissection protocols*".

[3] The notation $\Pr_{\mathcal{D}}[\mathcal{E}]$ denotes the probability of an event $\mathcal{E}$ under distribution $\mathcal{D}$.

Definition 7 (contiguous subset of inputs) *Given a permutation Π of $\{0,1\}^k$, let $\prec_\Pi$ denote the total order over $\{0,1\}^k$ that Π induces,* i.e., $\forall\, a,b \in \{0,1\}^k$, $a \prec_\Pi b$ *provided b comes after a in Π. Then, $I \subseteq \{0,1\}^k$* contiguous *with respect to Π if* $\forall\, a,b\in I,\ \forall\, c\in\{0,1\}^k\colon\ a \prec_\Pi c \prec_\Pi b \Longrightarrow c \in I$.

Definition 8 (dissection protocol) *Given a function $f : \{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}^t$ and permutations Π_1, Π_2 of $\{0,1\}^k$, a protocol for f is a* dissection protocol *with respect to (Π_1, Π_2) if, at each communication step, the maintained subset of inputs of each party_i is contiguous with respect to Π_i.*

Observe that *every* protocol P can be regarded as a dissection protocol with respect to *some* permutations over inputs by simply constructing the permutation so that it is consistent with the way P updates the maintained sets of inputs. However, *not* every protocol is a dissection protocol with respect to *specific* permutations. Consider, for example, the case that both Π_1 and Π_2 are the permutation over $\{0,1\}^k$ that orders the elements from lowest to highest binary values. Observe that a protocol that is a dissection protocol with respect to these permutations *cannot* ask questions of the form "Is your input odd or even?", for these questions partition the space of inputs into *non-contiguous* subsets with respect to (Π_1, Π_2).

A special case of interest of the dissection protocol is the "bisection type" protocols that have been investigated in the literature in many contexts [8, 10].

Definition 9 (bisection, c-bisection and bounded-bisection protocols) *For a constant $c \in \left[\frac{1}{2}, 1\right)$, a dissection protocol with respect to the permutations (Π_1, Π_2) is called a c-bisection protocol provided at each communication step each party_i partitions its input space of size z into two halves of size $c\,z$ and $(1-c)\,z$. A bisection protocol is simply a $\frac{1}{2}$-bisection protocol. For an integer valued function $g(k)$ such that $0 \le g(k) \le k$, bounded-bisection$_{g(k)}$ is the protocol that runs a bisection protocol with $g(k)$ bisection operations followed by a protocol (if necessary) in which each party_i repeatedly partitions its input space into two halves one of which is of size exactly one.*

Definition 10 (tiling and non-tiling functions) *A function $f : \{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}^t$ is called a tiling function with respect to two permutations (Π_1, Π_2) of $\{0,1\}^k$ if the monochromatic regions in $A_{\Pi_1,\Pi_2}(f)$ form a tiling, and the number of monochromatic regions in this tiling is denote by $\mathsf{r}_f\,(\Pi_1, \Pi_2)$. Conversely, f is a non-tiling function if f is not a tiling function with respect to every pair of permutations (Π_1, Π_2) of $\{0,1\}^k$.*

Note that a function f that is tiling function with respect to permutations (Π_1, Π_2) may not be a tiling function with respect to a different set of permutations (Π_1', Π_2'). Also, a function f can be a tiling function with respect to two distinct permutation pairs (Π_1, Π_2) and (Π_1', Π_2') with a different number of monochromatic regions. Thus, indeed we need Π_1 and Π_2 in the definition of tiling functions and r_f.

Extensions to Multi-party Computation For the multi-party computation model involving $d > 2$ parties, the d-ary tiling function f has a permutation Π_i of $\{0,1\}^k$ for each i^{th} argument of f (or, equivalently for each party_i). A dissection protocol is generalized to a "round robin" dissection protocol in the following manner. In one "mega" round of communications, parties communicate in a fixed order, say $\mathsf{party}_1, \mathsf{party}_2, \ldots, \mathsf{party}_d$, and the mega round is repeated if necessary. Any communication by any party is made available to *all* the other parties. Thus, each communication of the dissection protocol partitions a d-dimensional space by an appropriate *set* of $(d-1)$-dimensional hyperplanes, where the missing dimension in the hyperplane correspond to the index of the party communicating.

5 Two-party Dissection Protocol for Tiling Functions

5.1 Boolean Tiling Functions

Lemma 1 *Any Boolean tiling function* $f\colon \{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}$ *with respect to some two permutations* (Π_1, Π_2) *can be computed in a perfectly privacy-preserving manner by a dissection protocol with respect to* (Π_1, Π_2).

Remark 1. The claim of Lemma 1 is false if f outputs three values.

5.2 Average and Worst Case Par for Non-Boolean Tiling Functions

Let $f : \{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}^t$ be a given tiling function with respect to permutations (Π_1, Π_2). Neither the c-bisection nor the bounded-bisection protocol performs well in terms of average Par on arbitrary tiling functions. In this section, we show that *any* tiling function f admits a dissection protocol that has a *small constant* average case Par. Moreover, we show that this result *cannot* be extended to the case of worst-case Pars.

Constant Average-case Par for Non-Boolean Functions Let D_u denote the uniform distribution over all input pairs. We define the notion of a c-approximate uniform distribution $\mathsf{D}_u^{\sim c}$; note that $\mathsf{D}_u^{\sim 0} \equiv \mathsf{D}_u$.

Definition 11 (c-approximate uniform distribution) *A c-approximate uniform distribution* $\mathsf{D}_u^{\sim c}$ *is a distribution in which the probabilities of the input pairs are close to that for the uniform distribution as a linear function of c, namely* $\max_{(\mathbf{x},\mathbf{y}),\,(\mathbf{x}',\mathbf{y}')\in\{0,1\}^k\times\{0,1\}^k} \left| \Pr_{\mathsf{D}_u^{\sim c}}[\mathbf{x}\,\&\,\mathbf{y}] - \Pr_{\mathsf{D}_u^{\sim c}}[\mathbf{x}'\,\&\,\mathbf{y}'] \right| \le c\,2^{-2k}$.

Theorem 1
(a) *A tiling function f with respect to permutations (Π_1, Π_2) admits a dissection protocol P with respect to the same permutations (Π_1, Π_2) using at most* $4\,\mathsf{r}_f(\Pi_1, \Pi_2)$ *communication steps such that* $\alpha_{\mathsf{D}_u^{\sim c}} \le 4 + 4\,c$.

(b) *For all* $0 \le c < 9/8$, *there exists a tiling function* $f\colon \{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}^2$ *such that, for any two permutations (Π_1, Π_2) of $\{0,1\}^k$, every dissection protocol with respect to (Π_1, Π_2) using any number of communication steps has* $\alpha_{\mathsf{D}_u^{\sim c}} \ge (11/9) + (2/81)c$.

Proof. We only provide the proof of **(a)**; a proof of the other part can be found in the full version of the paper. Let $\mathcal{S} = \{S_1, S_2, \ldots, S_{\mathsf{r}_f}\}$ be the set of $\mathsf{r}_f = \mathsf{r}_f(\Pi_1, \Pi_2)$ ideal monochromatic rectangles in the tiling of f induced by the permutations (Π_1, Π_2) and consider a protocol P that is a dissection protocol with respect to (Π_1, Π_2). Suppose that the ideal monochromatic rectangle $S_i \in \mathcal{S}$ has y_i elements, and P partitions this rectangle into t_i rectangles $S_{i,1}, \ldots, S_{i,t_i}$ having $z_{i,1}, \ldots, z_{i,t_i}$ elements, respectively. Then, it follows that

$$\alpha_{\mathsf{D}_u} = \sum_{(x_1,x_2)\in\mathcal{U}\times\mathcal{U}} \Pr_{\mathsf{D}_u}[x_1 \,\&\, x_2]\, \frac{\left|R^I(x_1,x_2)\right|}{\left|R^P(x_1,x_2)\right|}$$
$$= \sum_{i=1}^{\mathsf{r}_f}\sum_{j=1}^{t_i}\sum_{(x_1,x_2)\in S_{i,j}} \Pr_{\mathsf{D}_u}[x_1 \,\&\, x_2]\, \frac{y_i}{z_{i,j}} = \sum_{i=1}^{\mathsf{r}_f}\sum_{j=1}^{t_i} \frac{y_i}{2^{2k}} = \sum_{i=1}^{\mathsf{r}_f} \frac{t_i\, y_i}{2^{2k}}$$

Similarly, it follows that
$\alpha_{\mathsf{D}_u^{\sim c}} \leq \sum_{i=1}^{\mathsf{r}_f}\sum_{j=1}^{t_i}\sum_{(x_1,x_2)\in S_{i,j}} \frac{1+c}{2^{2k}} \times \frac{y_i}{z_{i,j}} = \sum_{i=1}^{\mathsf{r}_f}\sum_{j=1}^{t_i} \frac{(1+c)\,y_i}{2^{2k}} = \sum_{i=1}^{\mathsf{r}_f} \frac{(1+c)\,t_i\,y_i}{2^{2k}}$.

A binary space partition (Bsp) for a collection of *disjoint* rectangles in the two-dimensional plane is defined as follows. The plane is divided into two parts by cutting rectangles with a line if necessary. The two resulting parts of the plane are divided recursively in a similar manner; the process continues until at most one fragment of the original rectangles remains in any part of the plane. This division process can be naturally represented as a binary tree (Bsp-tree) where a node represents a part of the plane and stores the cut that splits the plane into two parts that its two children represent and each leaf of the Bsp-tree represents the final partitioning of the plane by storing at most one fragment of an input rectangle. The *size* of a Bsp is the *number of leaves* in the Bsp-tree.

Fact 1 [4][4] *Assume that we have a set $\mathcal{S}$ of disjoint axis-parallel rectangles in the plane. Then, there is a* Bsp *of $\mathcal{S}$ such that every rectangle in $\mathcal{S}$ is partitioned into at most* 4 *rectangles.*

Consider the dissection protocol corresponding to the Bsp in Fact 1. Then, using $\max_i\{t_i\} \leq 4$ we get $\alpha_{\mathsf{D}_u^{\sim c}} \leq \sum_{i=1}^{\mathsf{r}_f} \frac{4\,(1+c)\,y_i}{2^{2k}} = 4\,(1+c)$. The number of communication steps in this protocol is the height of the Bsp-tree, *i.e.*, $\leq 4\mathsf{r}_f$.

Large Worst-case Par **for Non-Boolean Functions** Can one extend the results of the last section to show that every tiling function admits a dissection protocol that achieves a good Par *even in the worst case*? We answer this question in the negative by presenting a tiling function for which *every* dissection protocol has *large* worst-case Par.

Theorem 2 *Let $k > 0$ be an even integer. Then, there exists a tiling function $f : \{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}^3$ with respect to some two permutations (Π_1, Π_2) such that, for any two permutations Π_1' and Π_2' of $\{0,1\}^k$, every dissection protocol for f with respect to (Π_1', Π_2') has $\alpha_{\text{worst}} > 2^{k/2} - 1$.*

[4] The stronger bounds by Berman, DasGupta and Muthukrishnan [2] apply to *average* number of fragments only.

6 Extensions of the Basic Two-party Setup

6.1 Non-tiling Functions

A natural extension of the class of tiling functions involves relaxing the constraint that each monochromatic region *must* be a rectangle.

Definition 12 (δ-tiling function) *A function $f:\{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}^t$ is a δ-tiling function with respect to permutations (Π_1, Π_2) of $\{0,1\}^k$ if each maximal monochromatic region of $A_{\Pi_1,\Pi_2}(f)$ is an union of at most δ disjoint rectangles.*

Proposition 1 *For any δ-tiling function f with respect to (Π_1, Π_2) with r maximal monochromatic regions, there is a dissection protocol P with respect to (Π_1, Π_2) using at most $4r\delta$ communication steps such that $\alpha_{\mathrm{D}_{\tilde{u}}^{c}} \leq (4+4c)\,\delta$.*

6.2 Multi-party Computation

How good is the average PAR for a dissection protocol on a d-dimensional tiling function? For a general d, it is non-trivial to compute *precise* bounds because each party_i has her/his own permutation Π_i of the input, the tiles are boxes of *full* dimension and hyperplanes corresponding to each step of the dissection protocol is of dimension *exactly* $d-1$. Nonetheless, we show that the average PAR is very high for dissection protocols even for 3 parties and uniform distribution, thereby suggesting that this quantification of privacy may not provide good bounds for three or more parties.

Theorem 3 *There exists a tiling function $f:\{0,1\}^k \times \{0,1\}^k \times \{0,1\}^k \mapsto \{0,1\}^{3k}$ such that, for any three permutations Π_1, Π_2, Π_3 of $\{0,1\}^k$, every dissection protocol with respect to (Π_1, Π_2, Π_3) must have $\alpha_{\mathrm{D}_u} = \Omega\left(2^k\right)$.*

Proof. In the sequel, for convenience *we refer to* 3*-dimensional hyper-rectangles simply by rectangles* and refer to the arguments of function f via *decimal equivalent of the corresponding binary numbers*. The tiling function for this theorem is adopted from an example of the paper by Paterson and Yao [14, 15] with appropriate modifications. The three arguments of f are referred to as dimensions 1, 2 and 3, respectively. Define the *volume* of a rectangle $R = [x_1, x_1'] \times [x_2, x_2'] \times [x_3, x_3'] \subseteq \{0, 1, \ldots, 2^k - 1\}^3$ as $\mathsf{Volume}(R) = \max\{0, \Pi_{i=1}^{3}(x_i' - x_i + 1)\}$, and let $[*]$ denote the interval $\left[0, 2^k - 1\right]$. We provide the tiling for the function f:

- For each dimension, we have a set of $\Theta\left(2^{2k}\right)$ rectangles; we refer to these rectangles as *non-trivial* rectangles for this dimension.
  - For dimension 1, these rectangles are of the form $[*] \times [2y, 2y] \times [2z, 2z]$ for every integral value of $0 \leq 2y, 2z < 2^k$.
  - For dimension 2, these rectangles are of the form $[2x, 2x] \times [*] \times [2z + 1, 2z + 1]$ for every integral value of $0 \leq 2x, 2z + 1 < 2^k$.
  - For dimension 3, these rectangles are of the form $[2x + 1, 2x + 1] \times [2y + 1, 2y + 1] \times [*]$ for every integral value of $0 \leq 2x + 1, 2y + 1 < 2^k$.
- The remaining "trivial" rectangles are each of unit volume such that they together cover the remaining input space.

Let $\mathcal{S}_{\text{non-trivial}}$ be the set of all non-trivial rectangles. Observe that:

– Rectangles in $\mathcal{S}_{\text{non-trivial}}$ are mutually disjoint since any two of them do not intersect in at least one dimension.
– *Each* rectangle in $\mathcal{S}_{\text{non-trivial}}$ has a volume of 2^k and thus the sum of their volumes is $\Theta\left(2^{3k}\right)$.

It now follows that the number of monochromatic regions is $O\left(2^{3k}\right)$. Suppose that a dissection protocol partitions, for $i = 1, 2, \ldots, |\mathcal{S}_{\text{non-trivial}}|$, the i^{th} non-trivial rectangle $R_i \in \mathcal{S}_{\text{non-trivial}}$ into t_i rectangles $R_{i,1}, R_{i,2}, \ldots, R_{i,t_i}$. Then,

$$\alpha_{\mathsf{D}_u} \stackrel{\text{def}}{=} \sum_{\substack{(x,y,z) \in \\ \{0,1\}^k \times \{0,1\}^k \times \{0,1\}^k}} \Pr_{\mathsf{D}_u}[x \,\&\, y \,\&\, z] \, \frac{\left|R^I(x,y,z)\right|}{\left|R^P(x,y,z)\right|} \geq \sum_{i=1}^{|\mathcal{S}_{\text{non-trivial}}|} \sum_{j=1}^{t_i} \sum_{(x,y,z)\in R_{i,j}} \Pr_{\mathsf{D}_u}[x \,\&\, y \,\&\, z] \, \frac{\mathsf{Volume}\,(R_i)}{\mathsf{Volume}\,(R_{i,j})}$$

$$= \textstyle\sum_{i=1}^{|\mathcal{S}_{\text{non-trivial}}|} \sum_{j=1}^{t_i} \frac{2^k}{2^{3k}} = \sum_{i=1}^{|\mathcal{S}_{\text{non-trivial}}|} \left(t_i/2^{2k}\right)$$

Thus, it suffices to show that $\sum_{i=1}^{|\mathcal{S}_{\text{non-trivial}}|} t_i = \Omega\left(2^{3k}\right)$. Let $\mathcal{Q}$ be the set of maximal *monochromatic* rectangles produced the partitioning of the entire protocol. Consider the two entries $p_{x,y,z} = (2x+1, 2y, 2z+1)$ and $p'_{x,y,z} = (2x, 2y, 2z)$. Note that $p_{x,y,z}$ belongs to a trivial rectangle since their third, first and second coordinate does not lie within *any* non-trivial rectangle of dimension 1, 2 and 3, respectively, whereas $p'_{x,y,z}$ belongs to the non-trivial rectangle $[*] \times [2\times(8y), 2\times(8y)] \times [2\times(8z), 2\times(8z)]$ of dimension 1. Thus, $p_{x,y,z}$ and $p'_{x,y,z}$ cannot belong to the same rectangle in $\mathcal{Q}$. Let $T = \bigcup \left\{ \{p_{\,8x,8y,8z},\, p'_{\,8x,8y,8z}\} \,|\, 64 < 16x, 16y, 16z < 2^k - 64 \right\}$. Clearly, $|T| = \Theta\left(2^{3k}\right)$. For an entry (x_1, x_2, x_3), let its neighborhood be defined by the ball $\mathsf{Nbr}(x_1, x_2, x_3) = \left\{ (x'_1, x'_2, x'_3) \,|\, \forall i : |x_i - x'_i| \leq 4 \right\}$. Note that $\mathsf{Nbr}(p_{\,8x,8y,8z}) \cap \mathsf{Nbr}(p_{\,8x',8y',8z'}) = \emptyset$ provided $(x,y,z) \neq (x',y',z')$. Next, we show that, to ensure that the two entries $p_{\,8x,8y,8z}$ and $p'_{\,8x,8y,8z}$ are in two different rectangles in $\mathcal{Q}$, the protocol must produce an *additional* fragment of one of the non-trivial rectangles in the neighborhood $\mathsf{Nbr}(p_{\,8x,8y,8z})$; this would directly imply $\sum_i t_i = \Omega\left(2^{3k}\right)$.

Consider the step of the protocol *before* which $p_{\,8x,8y,8z}$ and $p'_{\,8x,8y,8z}$ were contained inside the same rectangle, namely a rectangle Q that includes the rectangle $[16x, 16x+1] \times [16y, 16y] \times [16z, 16z+1]$, but after which they are in two different rectangles $Q_1 = [a'_1, b'_1] \times [a'_2, b'_2] \times [a'_3, b'_3]$ and $Q_2 = [a''_1, b''_1] \times [a''_2, b''_2] \times [a''_3, b''_3]$. Remember that both Q_1 and Q_2 must have the same two dimensions and these two dimensions must be the same as the corresponding dimensions of Q. The following cases arise.

Case 1(split via the 1^{st} **coordinate):** $[a'_2, b'_2] = [a''_2, b''_2] \supseteq [16y, 16y]$, $[a'_3, b'_3] = [a''_3, b''_3] \supseteq [16z, 16z+1]$, $b'_1 = 16x$ and $a''_1 = 16x+1$. Then, a new fragment of a non-trivial rectangle of dimension 2 is produced at $[16x, 16y, 16z] \in \mathsf{Nbr}(p_{\,8x,8y,8z})$.

Case 2(split via the 2^{nd} **coordinate):** $[a'_1, b'_1] = [a''_1, b''_1] \supseteq [16x, 16x+1]$ and $[a'_3, b'_3] = [a''_3, b''_3] \supseteq [16z, 16z+1]$. This case is not possible.

Case 3(split via the 3^{rd} **coordinate):** $[a'_1, b'_1] = [a''_1, b''_1] \supseteq [16x, 16x+1]$, $[a'_2, b'_2] = [a''_2, b''_2] \supseteq [16y, 16y]$, $b'_3 = 16z$ and $a''_3 = 16z+1$. Then, a new fragment of a non-trivial rectangle of dimension 1 is produced at $[16x, 16y, 16z] \in \mathsf{Nbr}(p_{\,8x,8y,8z})$.

7 Analysis of the Bisection Protocol for Two Functions

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0,1\}^k$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \{0,1\}^k$. The functions that we consider are the following:

set-covering: $f_{\wedge,\vee}(\mathbf{x},\mathbf{y}) = \bigwedge_{i=1}^{n} (x_i \vee y_i)$. To interpret this as a set-covering function, suppose that the universe $\mathcal{U}$ consists of n elements $e_1, e_2, \ldots, e_n$ and the vectors $\mathbf{x}$ and $\mathbf{y}$ encode membership of the elements in two sets $S_{\mathbf{x}}$ and $S_{\mathbf{y}}$, *i.e.*, x_i (respectively, y_i) is 1 if and only if $e_i \in S_{\mathbf{x}}$ (respectively, $e_i \in S_{\mathbf{y}}$). Then, $f_{\wedge,\vee}(\mathbf{x},\mathbf{y}) = 1$ if and only if $S_{\mathbf{x}} \cup S_{\mathbf{y}} = \mathcal{U}$.

equality: $f_{=}(\mathbf{x},\mathbf{y}) = 1$ if $x_i = y_i$ for all $1 \leq i \leq k$, and $f_{=}(\mathbf{x},\mathbf{y}) = 0$ otherwise.

A summary of our bounds is as follows: for $f_{\wedge,\vee}$, $\alpha_{\text{worst}} \geq \alpha_{\mathsf{D}_u} \geq \left(\frac{3}{2}\right)^{2k}$; for $f_{=}$, $\alpha_{\mathsf{D}_u} = 2^k - 2 + 2^{1-k}$, and $\alpha_{\text{worst}} = 2^{2k-1} - 2^{k-1}$.

References

1. A. Ghosh, T. Roughgarden and M. Sundararajan. *Universally utility-maximizing privacy mechanisms*, 41st ACM Symp. on Theory of Computing, 351-360, 2009.
2. P. Berman, B. DasGupta and S. Muthukrishnan. *On the Exact Size of the Binary Space Partitioning of Sets of Isothetic Rectangles with Applications*, SIAM Journal of Discrete Mathematics, 15 (2), 252-267, 2002.
3. R. Bar-Yehuda, B. Chor, E. Kushilevitz and A. Orlitsky. *Privacy, additional information, and communication*, IEEE Trans. on Inform. Theory, 39, 55-65, 1993.
4. F. d'Amore and P. G. Franciosa. *On the optimal binary plane partition for sets of isothetic rectangles*, Information Processing Letters, 44, 255-259, 1992.
5. D. Chaum, C. Crépeau and I. Damgaard. *Multiparty, unconditionally secure protocols*, 22th ACM Symposium on Theory of Computing, 11-19, 1988.
6. B. Chor and E. Kushilevitz. *A zero-one law for boolean privacy*, SIAM Journal of Discrete Mathematics, 4, 36-47, 1991.
7. C. Dwork. *Differential privacy*, 33rd International Colloquium on Automata, Languages and Programming, 1-12, 2006.
8. J. Feigenbaum, A. Jaggard and M. Schapira. *Approximate Privacy: Foundations and Quantification*, ACM Conference on Electronic Commerce, 167-178, 2010.
9. E. Grigorievaa, P. J.-J. Heringsb, R. Müllera and D. Vermeulena. *The communication complexity of private value single-item auctions*, Operations Research Letters, 34, 491-498, 2006.
10. E. Grigorievaa, P. J.-J. Heringsb, R. Müllera and D. Vermeulena. *The private value single item bisection auction*, Economic Theory, 30, 107-118, 2007.
11. E. Kushilevitz and N. Nisan. *Communication Complexity*, Cambridge University Press, 1997.
12. D. Kifer and B.-R. Lin. *An Axiomatic View of Statistical Privacy and Utility*, to appear in Journal of Privacy and Confidentiality.
13. E. Kushilevitz. *Privacy and communication complexity*, SIAM Journal of Discrete Mathematics, 5 (2), 273-284, 1992.
14. M.Paterson and F.F. Yao. *Efficient binary space partitions for hidden-surface removal and solid modeling*, Discrete & Computational Geometry, 5(1), 485-503, 1990.
15. M. Paterson and F. F. Yao. *Optimal binary space partitions for orthogonal objects*, Journal of Algorithms, 13, 99-113, 1992.
16. A. C. Yao. *Some complexity questions related to distributive computing*, 11th ACM Symposium on Theory of Computing, 209-213, 1979.