Bisimilarity Distances for Approximate Differential Privacy

Dmitry Chistikov[1], Andrzej S. Murawski[2], and David Purser[1]

[1] Centre for Discrete Mathematics and its Applications (DIMAP) & Department of Computer Science, University of Warwick, UK
[2] Department of Computer Science, University of Oxford, UK

Abstract. Differential privacy is a widely studied notion of privacy for various models of computation. Technically, it is based on measuring differences between probability distributions. We study ϵ, δ-differential privacy in the setting of labelled Markov chains. While the exact differences relevant to ϵ, δ-differential privacy are not computable in this framework, we propose a computable bisimilarity distance that yields a sound technique for measuring δ, the parameter that quantifies deviation from pure differential privacy. We show this bisimilarity distance is always rational, the associated threshold problem is in **NP**, and the distance can be computed exactly with polynomially many calls to an **NP** oracle.

Keywords: Bisimilarity distances · Kantorovich metric · Differential privacy · Labelled Markov chains · Bisimulation · Analysis of probabilistic systems.

1 Introduction

Bisimilarity distances were introduced by [16, 17], as a metric analogue of classic probabilistic bisimulation [23], to overcome the problem that bisimilarity is too sensitive to minor changes in probabilities. Such robustness is highly desirable, because probabilistic automata arising in practice may often be based on approximate probability values, extracted or learnt from real world data.

In this paper, we study the computation of bisimilarity distances related to differential privacy. Differential privacy [18] is a security property that ensures that a small perturbation of the input leads to only a small perturbation in the output, so that observing the output makes it difficult to determine whether a particular piece of information was present in the input. A variant, ϵ-differential privacy, considers the ratio difference (rather than the absolute difference) between probabilities.

We will be concerned with the more general concept of ϵ, δ-differential privacy, also referred to as *approximate differential privacy*. The δ parameter allows one to assess to what degree ϵ-differential privacy ("pure differential privacy") was achieved. We will design a version of bisimilarity distance which will constitute a sound upper bound on δ, thus providing a reliable measure of security.

From a verification perspective, a natural question is how to analyse systems with respect to ϵ, δ-differential privacy. We carry out our investigations in the setting where the systems are *labelled Markov chains (LMC)*, abstractions of autonomous systems with probabilistic behaviour and under partial observability. States of an LMC $\mathcal{M}$ can be thought of as generating probability distributions on sets of traces, and these sets are taken to correspond to observable events. Let $\mathcal{M}$ be a system, and suppose s and s' are two states (configurations) of $\mathcal{M}$. Then we will say that s and s' satisfy ϵ, δ-differential privacy if the distributions on traces from these states are sufficiently close. We consider the following problem: given an LMC $\mathcal{M}$, states s and s', and a value of ϵ, determine δ such that s and s' satisfy ϵ, δ-differential privacy. Unfortunately, the smallest of such δ is not computable [22], which motivates our search for upper bounds.

In the spirit of generalised bisimilarity pseudometrics [12], our distance, denoted bd_α, is based on the Kantorovich-style lifting of distance between states to distance between distributions. However, because the underpinning distances in our case turn out not to be metrics, the setting does not quite fit into the standard picture, which presents a technical challenge. We discuss how the proposed distance may be computed, using techniques from linear programming, linear real arithmetic, and computational logic. Our first result is that the distance always takes on rational values of polynomial size with respect to the size of the LMC and the bit size of the probability values associated with transitions (Theorem 1).

This is then used to show that the associated threshold problem ("is bd_α upper-bounded by a given threshold value for two given states?") is in **NP** (Theorem 2). Note that the distance can be approximated to arbitrary precision by solving polynomially many instances of the threshold problem. Finally, we show that the distance can be computed exactly in polynomial time, given an **NP** oracle (Theorem 3). This places it in (the search version of) **NP**, leaving the possibility of polynomial-time computation open.

Related Work Chatzikokolakis et al. [12] have advocated the development of Kantorovich pseudometrics, instantiated with any metric distance function (rather than absolute value) in the context of differential privacy. They did not discuss the complexity of calculating such pseudometrics, but asked whether it was possible to extend their techniques to ϵ, δ-differential privacy. Our paper shows the extent to which this can be achieved; the technical obstacle that we face is that our distances are not metrics. To the best of our knowledge, no complexity results on differential privacy for Markov chains have previously appeared in the literature, and we are the first to address this gap.

The computation of the standard bisimilarity distances has been the topic of a long running line of research [7], starting with approximation [8]. The distance was eventually determined to be computable in polynomial time using the ellipsoid method to solve an implicit linear program of exponential size [14]. This technique turns out slow in practice and further techniques have been developed which are faster but do not have such strong complexity guarantees [2, 26]. Because of the two-sided nature of our distances, the main system of constraints

that we introduce in our work involves a maximum of two quantities. This nonlinearity at the core of the problem prevents us from relying on the ellipsoid method and explains the gap between our **NP** upper bound and the polynomial-time algorithms of [14].

Tschantz et al. [28] first studied differential privacy using a notion similar to bisimulation, which was extended to a more general class of bisimulation relations by Xu et al. [31]. Both consider only ϵ-differential privacy, i.e. ratio differences, but do not examine how these could be computed.

An alternative line of research by Barthe et al. [5] concerns formal mechanised proofs of differential privacy. Recently, that direction has been related to coupling proofs [4] – this still requires substantial effort to choose the coupling, although recent techniques have improved this [1]. We complement this line of research by taking an algorithmic verification-centred approach.

The remainder of the paper is arranged as follows. Section 2 introduces the basic setting of labelled Markov chains. In section 3, we discuss ϵ, δ-differential privacy and in section 4 we define our distance. Section 5 develops technical results on our extended case of Kantorovich lifting. These are subsequently used in section 6 to underpin techniques for computing the relevant distances.

2 Labelled Markov Chains

Given a finite set S, let $Dist(S)$ be the set of probability distributions on S.

Definition 1. *A* labelled Markov chain *(LMC) $\mathcal{M}$ is a tuple $\langle S, \Sigma, \mu, \ell \rangle$, where S is a finite set of states, Σ is a finite alphabet, $\mu : S \to Dist(S)$ is the transition function and $\ell : S \to \Sigma$ is the labelling function.*

Like in [2, 7, 14, 26], our definition features labelled states. Variations, such as transition labels, can be easily accommodated within the setting. We also assume that all transition probabilities are rational, represented as a pair of binary integers. The bit sizes of these integers form part of the bit size of the representation $|\mathcal{M}|$. We will often write μ_s for $\mu(s)$.

In what follows, we study probabilities associated with infinite sequences of labels generated by LMC's. We specify the relevant probability spaces next using standard measure theory [3, 6]. Let us start with the definition of cylinder sets.

Definition 2. *A subset $C \subseteq \Sigma^\omega$ is a* cylinder set *if there exists $u \in \Sigma^*$ such that C consists of all infinite sequences from Σ^ω whose prefix is u. We then write C_u to refer to C.*

Cylinder sets play a prominent role in measure theory in that their finite unions can be used as a generating family (an algebra) for the set $\mathcal{F}$ of measurable subsets of Σ^ω (the cylindric σ-algebra). What will be important for us is that any measure ν on $\mathcal{F}$ is uniquely determined by its values on cylinder sets. Next we show how to assign a measure ν_s on $\mathcal{F}$ to an arbitrary state of an LMC. We start with several auxiliary definitions.

Definition 3. *Given $\mathcal{M} = \langle S, \Sigma, \mu, \ell \rangle$, let $\mu^+ : S^+ \to [0,1]$ and $\ell^+ : S^+ \to \Sigma^+$ be the natural extensions of μ and ℓ to S^+, i.e. $\mu^+(s_0 \cdots s_k) = \prod_{i=0}^{k-1} \mu(s_i)(s_{i+1})$ and $\ell^+(s_0 \cdots s_k) = \ell(s_0) \cdots \ell(s_k)$, where $k \geq 0$ and $s_i \in S$ $(0 \leq i \leq k)$. Note that, for any $s \in S$, we have $\mu^+(s) = 1$. Given $s \in S$, let $Paths_s(\mathcal{M})$ be the subset of S^+ consisting of all sequences that start with s.*

Definition 4. *Let $\mathcal{M} = \langle S, \Sigma, \mu, \ell \rangle$ and $s \in S$. We define $\nu_s : \mathcal{F} \to [0,1]$ to be the unique measure on $\mathcal{F}$ such that for any cylinder C_u we have*

$$\nu_s(C_u) = \sum \{\, \mu^+(p) \mid p \in Paths_s(\mathcal{M}),\, \ell^+(p) = u \,\}.$$

Our aim will be to compare states of labelled Markov chains from the point of view of differential privacy. Note that two states s, s' can be viewed as indistinguishable if $\nu_s = \nu_{s'}$. If they are not indistinguishable then the difference between them can be quantified using the *total variation distance*, defined by $tv(\nu, \nu') = \sup_{E \in \mathcal{F}} |\nu(E) - \nu'(E)|$. Given $\mathcal{M} = \langle S, \Sigma, \mu, \ell \rangle$ and $s, s' \in S$, we shall write $tv(s, s')$ to refer to $tv(\nu_s, \nu_{s'})$.

Remark 1. $tv(s, s')$ turns out surprisingly difficult to compute: it is undecidable whether the distance is strictly greater than a given threshold, and the non-strict variant of the problem ("greater or equal") is not known to be decidable [22].

To measure probabilities relevant to differential privacy, we will need to study a more general variant tv_α of the above distance, which we introduce next.

3 Differential Privacy

Differential privacy is a mathematical guarantee of privacy due to Dwork et al [18]. It is a property similar to non-interference: the aim is to ensure that inputs which are related in some sense lead to very similar outputs. The notion requires that for two related states there only ever be a small change in output probabilities, and therefore discerning the two is difficult, which maintains the privacy of the states. Below we cast the definition in the setting of labelled Markov chains.

Definition 5. *Let $\mathcal{M} = \langle S, \Sigma, \mu, \ell \rangle$ be a labelled Markov chain and let $R \subseteq S \times S$ be a symmetric relation. Given $\epsilon \geq 0$ and $\delta \in [0,1]$, we say that $\mathcal{M}$ is ϵ, δ-differentially private (wrt R) if, for any $s, s' \in S$ such that $(s, s') \in R$, we have*

$$\nu_s(E) \leq e^\epsilon \cdot \nu_{s'}(E) + \delta$$

for any measurable set $E \in \mathcal{F}$.

Remark 2. Note that each state $s \in S$ can be viewed as defining a random variable X_s with outcomes from Σ^ω such that $P(X_s \in E) = \nu_s(E)$. Then the above can be rewritten as $P(X_s \in E) \leq e^\epsilon\, P(X_{s'} \in E) + \delta$, which matches the definition from [18], where one would consider $X_s, X_{s'}$ neighbouring in some natural sense.

The above formulation is often called *approximate differential privacy*. For $\delta = 0$, one talks about (pure) ϵ-*differential privacy*. Note that then the above definition boils down to measuring the ratio between the probabilities of possible outcomes. δ is thus an indicator of the extent to which ϵ-differential privacy holds for the given states. Intuitively, one could interpret ϵ, δ-differential privacy as "ϵ-differential privacy with probability at least $1 - \delta$" [29]. Our work is geared towards obtaining sound upper bounds on the value of δ for a given ϵ.

Remark 3. What it means for two states to be related (as specified by R) is to a large extent domain-specific. In general, R makes it possible to spell out which states should not appear too different and, consequently, should enjoy a quantitative amount of privacy. In the typical database scenario, one would relate database states that differ by just one person. In our case, we refer to states of a machine, for which we would like it to be indiscernible as to which was the start state (we assume the states are hidden and the traces are observable).

To rephrase the inequality underpinning differential privacy in a more succinct form, it will be convenient to work with the *skewed distance* Δ_α, first introduced by Barthe et al [5] in the context of Hoare logics and ϵ, δ-differential privacy.

Definition 6 (Skewed Distance). *For $\alpha \geq 1$, let $\Delta_\alpha : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be defined by $\Delta_\alpha(x, y) = \max\{x - \alpha y,\, y - \alpha x,\, 0\}$.*

Remark 4. It is easy to see that Δ_α is anti-monotone with respect to α. In particular, because $\alpha \geq 1$, we have $\Delta_\alpha(x, y) \leq \Delta_1(x, y) = |x - y|$. Observe that $\Delta_2(9, 3) = 9 - 2 \times 3 = 3$, $\Delta_2(9, 6) = 0$ and $\Delta_2(6, 3) = 0$. Note that $\Delta_2(x, y) = 0$ need not imply $x = y$, i.e. Δ_2 is not a metric. Note also that the triangle inequality may fail: $\Delta_2(9, 3) > \Delta_2(9, 6) + \Delta_2(6, 3)$, i.e. Δ_2 is not a pseudometric[3]. This will complicate our technical development, because we will not be able to use the framework of [12] directly.

The significance of the skewed distance will be seen shortly in Fact 1. We first introduce the skewed analogue of the total variation distance called tv_α, for which tv is a special case ($\alpha = 1$).

Definition 7. *Let $\alpha \geq 1$. Given two measures ν, ν' on $(\Sigma^\omega, \mathcal{F})$, let*

$$tv_\alpha(\nu, \nu') = \sup_{E \in \mathcal{F}} \Delta_\alpha(\nu(E), \nu'(E)).$$

Following the convention for tv, $tv_\alpha(s, s')$ will stand for $tv_\alpha(\nu_s, \nu_{s'})$. Fact 1 is an immediate corollary of Definitions 5, 6, and 7.

Fact 1. *$\mathcal{M}$ is ϵ, δ-differentially private wrt R if and only if, for all $s, s' \in S$ such that $(s, s') \in R$, we have $tv_\alpha(s, s') \leq \delta$, where $\alpha = e^\epsilon$.*

Some values of tv_α are readily known. For instance, the distance between any bisimilar states turns out to be zero.

[3] A pseudometric must satisfy $m(x, x) = 0$, $m(x, y) = m(y, x)$ and $m(x, z) \leq m(x, y) + m(y, z)$. For metrics, one additionally requires that $m(x, y) = 0$ should imply $x = y$.

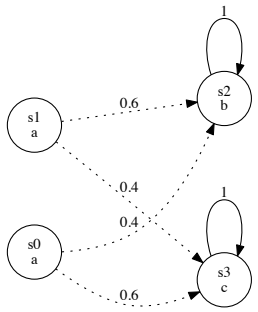

Fig. 1. States 1 and 2 are not bisimilar, but $tv_{1.5}(s_0, s_1) = 0$.

Definition 8. *A probabilistic bisimulation on an LMC $\mathcal{M} = \langle S, \Sigma, \mu, \ell \rangle$ is an equivalence relation $R \subseteq S \times S$ such that if $(s, s') \in R$ then $\ell(s) = \ell(s')$ and for all $X \in S/R$, $\sum_{u \in X} \mu(s)(u) = \sum_{u \in X} \mu(s')(u)$, i.e. related states have the same label and probability of transitioning into any given equivalence class.*

It is known that probabilistic bisimulations are closed under union and hence there exists a largest one, written $\sim$ and called *probabilistic bisimilarity*. Two states are called *bisimilar*, written $s \sim s'$, if $(s, s') \in \sim$. Equivalently, this means that the pair (s, s') belongs to a probabilistic bisimulation. It follows from [14, Proposition 9, Lemma 10], that for bisimilar s, s', we have $tv_1(s, s') = 0$. As $tv_\alpha(s, s') \leq tv_1(s, s')$ we obtain the following.

Lemma 1. *If $s \sim s'$ then $tv_\alpha(s, s') = 0$.*

In contrast to [12], the converse will not hold.

Example 1. In the LMC shown in Figure 1, states s_0 and s_1 are *not* bisimilar. To see this, observe first that s_2 must be the only state in its equivalence class with respect $\sim$, because other states have different labels. Now note that the probabilities of reaching s_2 from s_0 and s_1 respectively are different (0.4 vs 0.6).

However, for $\alpha = 1.5$, we have $tv_\alpha(s_0, s_1) = 0$, because $\Delta_\alpha(0.6, 0.4) = \max(0.6 - 1.5 \cdot 0.4, 0.4 - 1.5 \cdot 0.6, 0) = 0$.

In an "acyclic" system, tv_α can be calculated by exhaustive search: the natural algorithm is doubly exponential, as one needs to consider all possible events over all possible traces. However, in general, tv_α is not computable (Remark 1). Thus, in the remainder of the paper, we shall introduce and study another distance bd_α. It will turn out possible to compute it and it will provide a sound method for bounding δ for $\ln(\alpha), \delta$-differential privacy. Our main result will be Theorem 3: the new distance can be calculated in polynomial time, assuming an **NP** oracle. Pragmatically, this means that this new distance can be computed efficiently, assuming access to an appropriate satisfiability or theory solver.

4 Skewed Bisimilarity Distance

Our distance will be defined in the spirit of bisimilarity distances [12, 14, 16, 17] through a fixed point definition based on a variation of the Kantorovich lifting. To motivate its shape, let us discuss how one would go about calculating tv_α recursively. If $\ell(s) \neq \ell(s')$ then $\nu_s(C_{\ell(s)}) = 1$, $\nu_{s'}(C_{\ell(s)}) = 0$, therefore $tv_\alpha(s, s') = 1$. So, let us assume $\ell(s) = \ell(s')$. Given $E \subseteq \Sigma^\omega$ and $a \in \Sigma$, let $E_a = \{w \in \Sigma^\omega \mid aw \in E\}$. Then we have:

$$\begin{aligned} tv_\alpha(\nu_s, \nu_{s'}) &= \sup_{E \in \mathcal{F}} \Delta_\alpha(\nu_s(E), \nu_{s'}(E)) \\ &= \sup_{E_{\ell(s)} \in \mathcal{F}} \Delta_\alpha\big(\sum_{u \in S} \mu_s(u)\, \nu_u(E_{\ell(s)}), \sum_{u \in S} \mu_{s'}(u)\, \nu_u(E_{\ell(s)})\big). \end{aligned}$$

If we define $f : S \to [0, 1]$ by $f(u) = \nu_u(E_{\ell(s)})$, this can be rewritten as

$$\sup_{E_{\ell(s)} \in \mathcal{F}} \Delta_\alpha\big(\sum_{u \in S} \mu_s(u)\, f(u), \sum_{u \in S} \mu_{s'}(u)\, f(u)\big).$$

We have little knowledge of f, otherwise we could compute tv_α, but from the definition of tv_α, we do know that $\Delta_\alpha(f(v), f(v')) \leq tv_\alpha(v, v')$ for any $v, v' \in S$. Consequently, the following inequality holds.

$$tv_\alpha(s, s') \leq \sup_{\substack{f:S\to[0,1] \\ \forall v,v' \in S \Delta_\alpha(f(v),f(v')) \leq tv_\alpha(v,v')}} \Delta_\alpha\big(\sum_{u \in S} \mu_s(u) f(u), \sum_{u \in S} \mu_{s'}(u) f(u)\big)$$

The expression on the right is an instance of the Kantorovich lifting [15, 21], which uses ("lifts") the distance tv_α between states s, s' to define a distance between the distributions $\mu_s, \mu_{s'}$ associated with the states. We recall the definition of the Kantorovich distance between distributions in the discrete case, noting that then, for $\mu \in \mathit{Dist}(S)$, we have $\int f d\mu = \sum_{u \in S} f(u)\, \mu(u)$.

Definition 9 (Kantorovich). *Given $\mu, \mu' \in \mathit{Dist}(S)$ and a pseudometric $m : S \times S \to [0, 1]$, the* Kantorovich distance *between μ and μ' is defined to be*

$$K(m)(\mu, \mu') = \sup_{\substack{f:S\to[0,1] \\ \forall v,v' \in S |f(v)-f(v')| \leq m(v,v')}} \Big| \int f d\mu - \int f d\mu' \Big|.$$

Remark 5. The Kantorovich distance is also known under other names (e.g. Hutchinson, Wasserstein distance), having been rediscovered several times in history [15]. Chatzikokolakis et al. [12] studied the Kantorovich distance and related bisimulation distances when the absolute value distance above is replaced with another metric. For our purposes, instead of $|...|$, we need to consider Δ_α, even though Δ_α is not a metric and m may not be a pseudometric.

Definition 10 (Skewed Kantorovich). *Given $\mu, \mu' \in \mathit{Dist}(S)$ and a symmetric distance $d : S \times S \to [0, 1]$, the* skewed Kantorovich distance *between μ and μ' is defined to be*

$$K_\alpha(d)(\mu, \mu') = \sup_{\substack{f:S\to[0,1] \\ \forall v,v' \in S \ \Delta_\alpha(f(v),f(v')) \leq d(v,v')}} \Delta_\alpha\Big(\int f d\mu, \int f d\mu'\Big)$$

Note that setting $\alpha = 1$ gives the standard Kantorovich distance (Definition 9). Below we define a function operator, which will be used to define our distance.

Definition 11. *Let* $\Gamma_\alpha : [0,1]^{S\times S} \to [0,1]^{S\times S}$ *be defined as follows.*

$$\Gamma_\alpha(d)(s,s') = \begin{cases} K_\alpha(d)(\mu_s, \mu_{s'}) & \ell(s) = \ell(t) \\ 1 & \ell(s) \neq \ell(t) \end{cases}$$

Note that $[0,1]^{S\times S}$ equipped with the pointwise order, written $\sqsubseteq$, is a complete lattice and that Γ_α is monotone with respect that order (larger d permit more functions, thus larger supremum). Consequently, Γ_α has a least fixed point [27]. We take our distance to be exactly that point.

Definition 12 (Skewed Bisimilarity Distance). *Let* $bd_\alpha : S \times S \to [0,1]$ *be the least fixed point of* Γ_α.

Remark 6. Recall that the least fixed point is equal to the least pre-fixed point ($\min\{d \mid \Gamma_\alpha(d) \sqsubseteq d\}$).

Recall our initial remarks about the Kantorovich distance $K_\alpha(tv_\alpha)(\mu_s, \mu_{s'})$ over-approximating $tv_\alpha(s,s')$. They can be summarised by $tv_\alpha \sqsubseteq K_\alpha(tv_\alpha)$, i.e. tv_α is a post-fixed point of K_α. Since we want to bound tv_α as closely as possible, we can show that the least fixed point bd_α also bounds tv_α from above.

Lemma 2. $tv_\alpha \sqsubseteq bd_\alpha$.

Remark 7. The lemma is an analogue of Theorem 2 [12]. Its proof in [30] relied on the fact that the counterpart of Δ_α was a metric, which is not true in our case (unless $\alpha = 1$).

Just like Δ_α is anti-monotone with respect to α, so is bd_α. This means that $bd_\alpha \sqsubseteq bd_1$. The definition of bd_1 coincides with the definition of the classic bisimilarity pseudometric $\mathtt{d}_1$ (see e.g. [14]), which satisfies $\mathtt{d}_1(s,s') = 0$ if and only if s and s' are bisimilar. Consequently, we obtain the following corollary.

Corollary 1. *For any* $\alpha \geq 1$, *if* $s \sim s'$ *then* $bd_\alpha(s,s') = 0$.

As in the case of tv_α, we do not have the converse in our setting. Example 1 shows that $s_0 \not\sim s_1$ but we observe that $bd_{1.5}(s_0,s_1) = 0$. Observe:

$$\begin{aligned} bd_{1.5}(s_0,s_1) \quad &\leq \\ &\max_f \Big(\sum_{s\in S} f(s)(\mu_{s_0}(s) - 1.5\cdot\mu_{s_1}(s)), \sum_{s\in S} f(s)(\mu_{s_1}(s) - 1.5\cdot\mu_{s_0}(s)) \Big) \\ &\qquad = \max_f (f(s_2)(0.6 - 1.5\cdot 0.4) + f(s_3)(0.4 - 1.5\cdot 0.6), \\ &\qquad\qquad\qquad f(s_2)(0.4 - 1.5\cdot 0.6) + f(s_3)(0.6 - 1.5\cdot 0.4)). \end{aligned}$$

Notice the coefficients of $f(s)$ are all non-positive. Consequently, regardless of the restrictions on f, the maximising allocation will be $f(s) = 0$ and, thus, $bd_{1.5}(s_0,s_1) = 0$.

```
1  diningCrypto(payingCryptographer):
2     firstFlip  = flip(p, 1−p)
3    previousFlip = firstFlip
4    for cryptographer = 0 → n−1:
5      if cryptographer == n−1:
6        thisFlip = firstFlip
7      else :
8        thisFlip = flip(p, 1−p)
9      if (cryptographer == payingCryptographer):
10       announce(previousFlip == thisFlip)
11     else :
12       announce(previousFlip != thisFlip)
13     previousFlip = thisFlip
```

Fig. 2. Simulation of Dining Cryptographers Protocol

Example: Dining Cryptographers

In the dining cryptographer model [13], a ring of diners want to determine whether one of the diners paid or an outside body. If a diner paid, we do not want to reveal which of them it was. The protocol proceeds with each adjacent pair privately flipping a coin, each diner then reports the XOR of the two coin flips they observe, however if the diner paid he would report the negation of this. We can determine if one of them paid by taking the XOR of the announcements. With perfectly fair coins, the protocol guarantees privacy of the paying diner, but it is still differentially private if the coins are biased. If an outside body paid, there is no privacy to maintain so we only simulate the scenarios in which one of the diners did pay. The scenario where Cryptographer 0 paid must have similar output distribution to Cryptographer 1 paying, so that it can be determined that one of them did pay, but not which. The internal configuration of the machine is always assumed to be hidden, but the announcements are made public whilst maintaining the privacy of the participating Cryptographer (and the internal states).

The LMC in Figure 3 shows the 2-person dining cryptographers protocol (Figure 2) starting from Cryptographers 0 and 1 using weighted coins with $p = \frac{49}{100}$. The states of the machine encode the 5 variables that need to be tracked. To achieve ϵ, δ-differential privacy with $\alpha = e^{\epsilon} = 1.0002$ the minimal (true) value of δ is 0.00030004. Our methods generate a correct upper bound $bd_{\alpha}(s_0, s_1) = 0.0004$, showing $\ln(1.0002), 0.0004$-differential privacy. The protocol could be played with n players, requiring $O(n^2)$ states, for all possible assignments of paying cryptographer and current cryptographer. In a two-person scenario, the diners would know which of them had paid but an external observer of the output would only learn that one of them paid, not which.

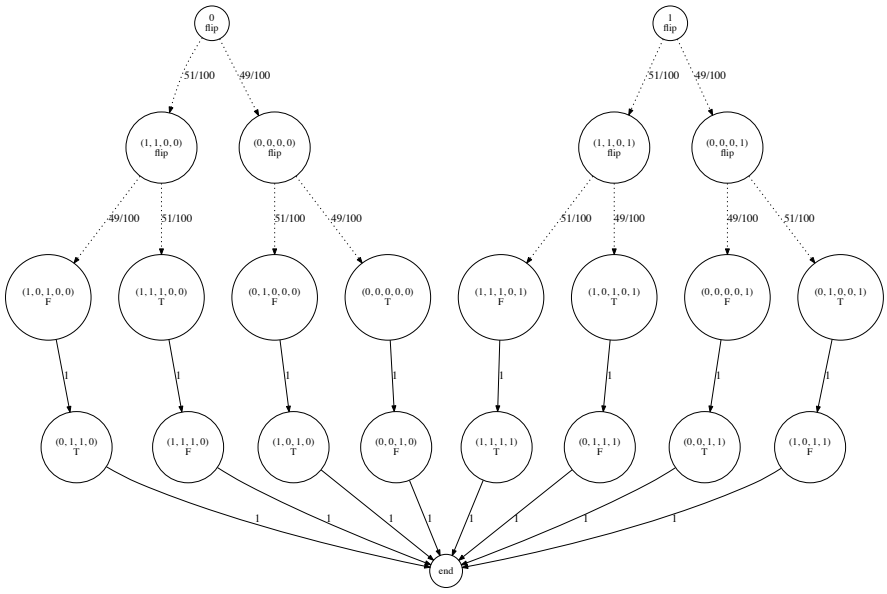

Fig. 3. Markov Chain for 2 dining cryptographers: state 0 (resp. 1) denotes Cryptographer 0 (resp. 1) paid. The first line of a node is the state name, the second line is the label of the state.

5 Skewed Kantorovich distances

Here we discuss how to calculate our variant of the Kantorovich distance. This will inform the next section, in which we look into computing bd_α.

Recall the definition of $K_\alpha(d)(\mu, \mu')$ from Definition 10. In the general case of $\Delta_\alpha(a, b)$, both $a - \alpha b$ and $b - \alpha a$ could be negative, so the maximum with 0 is taken. However, within the Kantorovich function, the constant function $f(i) = 0$ is a valid assignment, which achieves 0 in either case ($0 - \alpha \times 0 = 0$). Consequently, we can simplify the definition of Δ_α to omit the 0 case inside K_α.

If $\alpha = 1$ then Δ_α is the absolute value function and it is known that the distance corresponds to a single instance of a linear programming problem [9]. However, this is no longer true in our case due to the shape of $\Delta_\alpha(x, y) = \max(x - \alpha y, y - \alpha x)$. Still, one can present the calculation as taking the maximum of a pair of linear programs. We shall refer to this formulation as the "primal form" of $K_\alpha(d)$. We give the first program below, the other is its symmetric variant with μ, μ' reversed. Below we write f_i for $f(i)$ and let i, j range over S, and assume that d is symmetric.

$$\max_{f \in [0,1]^S} \Big(\sum_i f_i \mu(i) - \alpha \sum_i f_i \mu'(i) \Big) \qquad \text{subject to} \qquad \forall i, j \quad f_i - \alpha f_j \le d_{i,j}$$

The standard Kantorovich distance ($\alpha = 1$) is often presented in the following dual form when m is a pseudometric, based on the minimum coupling between the two distributions μ and μ', weighted by the distance function.

$$K(m)(\mu,\mu') = \min_{\omega\in[0,1]^{S\times S}} \sum_{i,j} \omega_{i,j} m_{i,j} \qquad \text{subject to} \qquad \begin{array}{ll} \forall i & \sum_j \omega_{i,j} = \mu(i) \\ \forall j & \sum_i \omega_{i,j} = \mu'(j) \end{array}$$

Remark 8. The dual form can be viewed as an optimal transportation problem in which an arbitrarily divisible cargo must be transferred from one set of locations (represented by a copy S^L of S) to another (represented by a different copy S^R of S). Each state $s^R \in S^R$ must receive $\mu(s)$, while each state $s^L \in S^L$ must send $\mu'(s)$. If $\omega_{i,j}$ is taken to represent the amount that gets sent from j^L to i^R then the above conditions restrict ω in accordance with the sending and receiving budgets. If $d_{i,j}$ represents the cost of sending from j^L to i^R then the objective function $\sum_{i,j} \omega_{i,j} \cdot d_{i,j}$ corresponds to the overall cost of transport. Consequently, the problem is referred to as a mass transportation problem [21].

To achieve a similar "dual form" in our case, we take the dual form of each of our linear programs. Then we can calculate the distance by taking the maximum of the two minima. The shape of the dual is given below on the right.

Lemma 3.

$$\begin{array}{rcl} \max\limits_{f\in[0,1]^S} \big(\sum\limits_i f_i\mu(i) - \alpha\sum\limits_i f_i\mu'(i)\big) & = & \min\limits_{\omega\in[0,1]^{S\times S},\tau,\gamma\in[0,1]^S} \sum\limits_{i,j} \omega_{i,j}\cdot d_{i,j} \\ \text{subject to} & & \text{subject to} \\ \forall i,j\ f_i - \alpha f_j \le d_{i,j} & & \forall i : \sum_j \omega_{i,j} + \tau_i - \gamma_i = \mu(i) \\ & & \forall j : \sum_i \omega_{i,j} + \frac{\tau_j-\gamma_j}{\alpha} \le \mu'(j) \end{array}$$

The dual form presented above is a simplified (but equivalent) form of the immediate dual obtained via the standard LP recipe. Note that the polytope we are optimising over is independent of d, which appears only in the objective function. The dual of the other linear program is obtained by swapping μ, μ'.

Remark 9. In the skewed case, we optimise over the following polytope

$$\Omega_{\mu,\mu'} = \left\{ \omega \in [0,1]^{S\times S} \quad | \quad \exists \gamma,\tau \in [0,1]^S \quad \begin{array}{l} \forall i : \sum_j \omega_{i,j} + \tau_i - \gamma_i = \mu(i) \\ \forall j : \sum_i \omega_{i,j} + \frac{\tau_j-\gamma_j}{\alpha} \le \mu'(j) \end{array} \right\}$$

One can also view it as a kind of transportation problem. As before, cargo can be transferred through the standard routes with ω at a cost d, but there are additional, cost-free routes between corresponding pairs s^L and s^R (represented by τ_s) and back (represented by γ_s). These extra routes are quite peculiar. En route from s^L to s^R the cargo 'grows': when $\frac{\tau_s}{\alpha}$ is sent from s^L, a larger amount of τ_s is received at s^R. Overall, the total amount of cargo sent may be less than that received, so the sending constraints are now inequalities. From s^R to s^L the cargo 'shrinks': when γ_s is sent from s^R, only $\frac{\gamma_s}{\alpha}$ is received by s^L.

It is immediate that τ routes can be useful. The γ routes may be useful for optimisation under two conditions. Firstly the shrinkage of the cargo must be made up elsewhere, i.e., through 'growing' τ routes. Additionally the cost $\alpha \times d(s_1^L, s^R) + d(s^L, s_2^R)$ is lower than $d(s_1^L, s_2^R)$, which may well be the case due to the lack of triangle inequality. Note that it is not possible to satisfy the receiving constraints if, in total, more is sent through γ routes than received through τ routes, so $\sum_s (\tau_s - \gamma_s) \geq 0$. Therefore, the vector ω (the coupling) may be smaller than its equivalent in the standard Kantorovich case.

We arrive at the following formulation, which we call the "dual form":

$$K_\alpha(d)(\mu, \mu') = \max \left\{ \min_{\omega \in \Omega_{\mu,\mu'}} \sum_{i,j} \omega_{i,j} \cdot d_{i,j},\ \min_{\omega \in \Omega_{\mu',\mu}} \sum_{i,j} \omega_{i,j} \cdot d_{i,j} \right\}.$$

Note that $K_\alpha(d)(\mu, \mu')$ can be computed in polynomial time as a pair of linear programs in either primal or dual form, and taking the maximum (in either case). In our calculations related to bd_α, the distributions μ, μ' will always be taken to be $\mu_s, \mu_{s'}$ respectively, for some $s, s' \in S$. The ability to switch between primal and dual form will play a useful role in our complexity-theoretic arguments.

6 Computing bd_α

We start off by observing that all distances $bd_\alpha(s, s')$ are rational and can be expressed in polynomial size with respect to $\mathcal{M}$. To that end, we exploit a result by Sontag [25], which states that, without affecting satisfiability, quantification in the first-order fragment of linear real arithmetic (LRA) can be restricted to rationals of polynomial size with respect to formula length (as long as all coefficients present in the formula are rational). Consequently, if we can express "there exists a least fixed point d of Γ_α" in this fragment (with a polynomial increase in size), we can draw the intended conclusion.

We give the relevant formula in Figure 4. The formula asserts the existence of a distance d, which is a pre-fixed point of Γ_α ($\forall f.\phi(d, f)$) such that any other pre-fixed point d' of Γ_α is greater. Note that $\forall f.\phi(f, d)$ exploits the fact that $\max_f A(f) \leq d(s, s')$ is equivalent to $\forall f(A(f) \leq d(s, s'))$. Sontag's result then implies the following.

Theorem 1. *Values of bd_α are rational. There exists a polynomial p such that for any LMC $\mathcal{M}$ and $s, s' \in S$, the size of bd_α (in binary) can be bounded from above by a polynomial in $|\mathcal{M}|$.*

Remark 10. Sontag [25] uses the fact mentioned above to relate the alternation hierarchy within LRA to the polynomial hierarchy **PH**: formulae of the form $\exists x_1 \forall x_2 \ldots Q x_k F(x_1 \ldots x_k)$ (with quantifier-free F) correspond to $\Sigma_k^{\mathbf{P}}$ (and formulae starting with $\forall$ to $\Pi_k^{\mathbf{P}}$). Recall that $\Sigma_1^{\mathbf{P}} = \mathbf{NP}$.

Next we focus on the following decision problem for bd_α.

$$\exists d \in [0,1]^{S\times S}\ (\forall f \in [0,1]^S \phi(d,f) \wedge \forall d' \in [0,1]^{S\times S} \forall f \in [0,1]^S (\phi(d',f) \implies \bigwedge_i d_i \leq d'_i))$$

$$\phi(d,f) = \bigwedge_{s,s'} \begin{cases} d_{s,s'} = 1 & \ell(s) \neq \ell(s') \\ (\bigwedge_{i,j} f_i - \alpha f_j \leq d_{i,j} \wedge f_j - \alpha f_i \leq d_{i,j}) & \ell(s) = \ell(s') \\ \quad \implies (\sum_i f_i \mu_s(i) - \alpha \sum_i f_i \mu_{s'}(i) \leq d_{s,s'} & \\ \qquad \wedge \sum_i f_i \mu_{s'}(i) - \alpha \sum_i f_i \mu_s(i) \leq d_{s,s'}) & \end{cases}$$

Fig. 4. Logical formulation of least pre-fixed point.

BD-THRESHOLD: given $s, s' \in S$ and $\theta \in \mathbb{Q}$, is it the case that $bd_\alpha(s,s') \leq \theta$?

Recall that the analogous problem for tv_α is undecidable (Remark 1). In our case, the problem turns out to be decidable and the argument does not depend on whether $<$ or $\leq$ is used. To establish decidability we can observe that $bd_\alpha(s,s') \leq \theta$ can be expressed in LRA simply by adding $d(s,s') \leq \theta$ to the formula from Figure 4. By Sontag's results, this not only yields decidability but also membership in $\Sigma_2^{\mathbf{P}}$. Recall that $\mathbf{NP} \subseteq \Sigma_2^{\mathbf{P}} \subseteq \mathbf{PH} \subseteq \mathbf{PSPACE}$.

We can simplify the formula, though, using $bd_\alpha = \min\{d \mid \Gamma_\alpha(d) \sqsubseteq d\}$. Then $bd_\alpha(s,s') \leq \theta$ can be specified as the existence of a pre-fixed point d such that $d(s,s') \leq \theta$. This can be done as follows, using $\phi(d,f)$ from Figure 4.

$$\exists d \in [0,1]^{S\times S}\ (\ \forall f \in [0,1]^S \phi(d,f)\ \wedge\ d(s,s') \leq \theta\ )$$

Note that the universal quantification over f remains, i.e. we can still only conclude that the problem is in $\Sigma_2^{\mathbf{P}}$. To overcome this, we shall use the dual form instead (Lemma 3). This will enable us to eliminate the universal quantification and replace it with existential quantifiers using the fact that $\min_\omega A(\omega) \leq B$ is equivalent to $\exists \omega (A(\omega) \leq B)$. The resultant formula is shown in Figure 5.

Note the formula is not linear due to $\omega_{i,j} \cdot d_{i,j}$. However, because we know (Theorem 1) that bd_α corresponds to an assignment of poly-sized rationals, we can consider the formula with d fixed at bd_α. Then it does become an LRA formula (of polynomially bounded length with respect to $|\mathcal{M}|$) and we can again conclude that the assignments of ω, γ, τ must also involve rationals whose size is polynomially bounded. Consequently, the formula implies membership of our problem in $\Sigma_1^{\mathbf{P}} = \mathbf{NP}$: it suffices to guess the satisfying assignment, guaranteed to be rational and of polynomial size.

Theorem 2. BD-THRESHOLD *is in* **NP**.

The decidability of BD-THRESHOLD makes it possible to approximate $bd_\alpha(s,s')$ to arbitrary (rational) precision ϵ by binary search. This will involve $O(|\epsilon|)$ calls to the oracle for BD-THRESHOLD (where $|\epsilon|$ is the number of bits required to represent ϵ in binary).

What's more, assuming the oracle, one can actually find the exact value of $bd_\alpha(s,s')$ in polynomial time (wrt $\mathcal{M}$). This exploits the fact that the value of

$$\text{BD-THRESHOLD}(s,s',\theta) = \exists (d_{i,j})_{i,j\in S} \quad \bigwedge_{i,j\in S} (0 \le d_{i.j} \le 1) \ \wedge \ \mathit{prefixed}(d) \ \wedge \ d_{s,s'} \le \theta$$

$$\mathit{prefixed}(d) = \bigwedge_{q,q'\in S} \begin{cases} d_{q,q'} = 1 & \ell(q) \neq \ell(t) \\ \mathit{prefixed}_1(d, d_{q,q'}, q, q') & \ell(q) = \ell(t) \\ \quad \wedge \ \mathit{prefixed}_1(d, d_{q,q'}, q', q) & \end{cases}$$

$$\mathit{prefixed}_1(d,x,q,q') = \exists (\omega_{i,j})_{i,j\in S} \quad \exists (\gamma_i)_{i\in S} \quad \exists (\tau_i)_{i\in S} \quad \sum_{i,j\in S} \omega_{i,j} \cdot d_{i,j} \le x$$
$$\wedge \bigwedge_{i,j\in S} (0 \le \omega_{i,j} \le 1) \wedge \bigwedge_{i\in S} (0 \le \gamma_i \le 1 \ \wedge \ 0 \le \tau_i \le 1)$$
$$\wedge \bigwedge_{i\in S} (\sum_{j\in S} \omega_{i,j} - \gamma_i + \tau_i = \mu_q(i)) \wedge \bigwedge_{j\in S} (\sum_{i\in S} \omega_{i,j} + \frac{\tau_j - \gamma_j}{\alpha} \le \mu_{q'}(j))$$

Fig. 5. NP Formula for BD-THRESHOLD

bd_α is rational and its size is polynomially bounded, so one can find it by approximation to a carefully chosen level of precision and then finding the relevant rational with the continued fraction algorithm [19, 20].

Theorem 3. *bd_α can be calculated in polynomial time with an* **NP** *oracle.*

As a consequence, the problem of computing bd_α reduces to propositional satisfiability, i.e., can be encoded in SAT. This justifies, for instance, the following approach: treat every variable as a ratio of two integers from an exponential range, and give the system of resulting constraints to an Integer Arithmetic or SAT solver. While this might look like resorting to a general-purpose "hammer", Theorem 3 is necessary for this method to work: it is not, in fact, possible to solve general polynomial constraint systems relying just on SAT.[4]

We expect, however, this direct approach to be inferior to the following observation. Theorem 1 reveals that the variables in our constraint system need not assume irrational values or have large bit representations. Thus, one can give the system to a more powerful theory solver, or an optimisation tool, but to expect that the existence of simple and small models (solutions) will help the SMT heuristics (resp. optimization engines) to find them quickly.

7 Conclusion and Further Work

We have demonstrated that bisimilarity distances can be used to determine differential privacy parameters, despite their non-metric properties. We have

[4] More precisely, the existence of such a procedure would be a breakthrough in the computational complexity theory, showing that **NP** = $\exists\mathbb{R}$. This would imply that a multitude of problems in computational geometry could be solved using SAT solvers [11, 24]. Unlike for bd_α, variable assignments in these problems may need to be irrational, even if all numbers in the input data are integer or rational.

established that the complexity of finding these values is polynomial, relative to an **NP** oracle. Yet, it may still be possible to obtain a polynomial algorithm—although much like in the case of the classical bisimilarity distances and linear programming, it may not necessarily outperform theoretically slower procedures.

We conjecture that bd_α, which we defined as the least fixed point of the operator Γ_α, may in fact be characterized as the unique fixed point of a similar operator. By the results of Etessami and Yannakakis [19], it would then follow that bd_α can be computed in **PPAD**, a smaller complexity class, improving upon our **NP** upper bound and matching the complexity of a closely related setting (see below). The reason is the continuity of Γ_α, which follows from the properties of the polytope over which f ranges (in the definition of $K_\alpha(d)$). Whether bd_α can in fact be computed in polynomial time or is **PPAD**-hard seems to be a challenging open question.

Our existing work is limited to labelled Markov chains, or fully probabilistic automata. However, the standard bisimulation distances can also be defined on deterministic systems, where their computational complexity is **PPAD** [10]. In our scenario, the privacy can only be analysed between two start states, but it is also reasonable to allow an input in the form of a trace or sequence of actions; the output would also be a trace. Here the choice of labels (at a specific state) would correspond to decisions taken by the user, and the availability of only one label would mean that this is the output. This setting would support a broader range of scenarios that could be modelled and verified as differentially private.

Acknowledgement David Purser gratefully acknowledges funding by the UK Engineering and Physical Sciences Research Council (EP/L016400/1), the EPSRC Centre for Doctoral Training in Urban Science. Andrzej Murawski is supported by a Royal Society Leverhulme Trust Senior Research Fellowship and the International Exchanges Scheme (IE161701).

References

1. Albarghouthi, A., Hsu, J.: Synthesizing coupling proofs of differential privacy. Proceedings of the ACM on Programming Languages **2**, 58:1–58:30 (2018)
2. Bacci, G., Bacci, G., Larsen, K.G., Mardare, R.: On-the-fly exact computation of bisimilarity distances. In: TACAS. pp. 1–15. Springer (2013)
3. Baier, C., Katoen, J.P.: Principles of model checking. MIT Press (2008)
4. Barthe, G., Espitau, T., Grégoire, B., Hsu, J., Stefanesco, L., Strub, P.Y.: Relational reasoning via probabilistic coupling. In: LPAR. pp. 387–401. Springer (2015)
5. Barthe, G., Köpf, B., Olmedo, F., Zanella Béguelin, S.: Probabilistic relational reasoning for differential privacy. In: POPL. pp. 97–110. ACM (2012)
6. Billingsley, P.: Probability and Measure. John Wiley and Sons, 2nd edn. (1986)
7. van Breugel, F.: Probabilistic bisimilarity distances. ACM SIGLOG News **4**(4), 33–51 (2017)
8. van Breugel, F., Sharma, B., Worrell, J.: Approximating a behavioural pseudometric without discount. In: FoSSaCS. pp. 123–137. Springer (2007)

9. van Breugel, F., Worrell, J.: An algorithm for quantitative verification of probabilistic transition systems. In: CONCUR. pp. 336–350. Springer (2001)
10. van Breugel, F., Worrell, J.: The complexity of computing a bisimilarity pseudometric on probabilistic automata. In: Horizons of the Mind. A Tribute to Prakash Panangaden, LNCS, vol. 8464, pp. 191–213. Springer (2014)
11. Cardinal, J.: Comput. geometry column 62. SIGACT News **46**(4), 69–78 (2015)
12. Chatzikokolakis, K., Gebler, D., Palamidessi, C., Xu, L.: Generalized bisimulation metrics. In: CONCUR. pp. 32–46. Springer (2014)
13. Chaum, D.: The dining cryptographers problem: Unconditional sender and recipient untraceability. J. Cryptology **1**(1), 65–75 (1988)
14. Chen, D., van Breugel, F., Worrell, J.: On the complexity of computing probabilistic bisimilarity. In: FoSSaCS. pp. 437–451. Springer (2012)
15. Deng, Y., Du, W.: The Kantorovich metric in computer science: A brief survey. Electronic Notes in Theoretical Computer Science **253**(3), 73–82 (2009)
16. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Metrics for labelled markov processes. Theoretical computer science **318**(3), 323–354 (2004)
17. Desharnais, J., Jagadeesan, R., Gupta, V., Panangaden, P.: The metric analogue of weak bisimulation for probabilistic processes. In: LICS. pp. 413–422. IEEE (2002)
18. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: TCC. pp. 265–284. Springer (2006)
19. Etessami, K., Yannakakis, M.: On the complexity of Nash equilibria and other fixed points. SIAM J. Comput. **39**(6), 2531–2597 (2010)
20. Grötschel, M., Lovász, L., Schrijver, A.: Geometric Algorithms and Combinatorial Optimization, Algorithms and Combinatorics, vol. 2. Springer (1988)
21. Kantorovich, L.V.: On the translocation of masses. Doklady Akademii Nauk SSSR **37(7-8)**, 227–229 (1942)
22. Kiefer, S.: On computing the total variation distance of hidden markov models. In: ICALP, pp. 130:1–130:13 (2018)
23. Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing. Information and computation **94**(1), 1–28 (1991)
24. Schaefer, M., Stefankovic, D.: Fixed points, Nash equilibria, and the existential theory of the reals. Theory Comput. Syst. **60**(2), 172–193 (2017)
25. Sontag, E.D.: Real addition and the polynomial hierarchy. IPL **20**(3), 115–120 (1985)
26. Tang, Q., van Breugel, F.: Computing probabilistic bisimilarity distances via policy iteration. In: CONCUR. pp. 22:1–22:15. Leibniz-Zentrum (2016)
27. Tarski, A.: A lattice-theoretical fixpoint theorem and its applications. Pacific Journal of Mathematics **5**(2), 285–309 (1955)
28. Tschantz, M.C., Kaynar, D., Datta, A.: Formal verification of differential privacy for interactive systems. ENTCS **276**, 61–79 (2011)
29. Vadhan, S.P.: The complexity of differential privacy. In: Tutorials on the Foundations of Cryptography, pp. 347–450. Springer (2017)
30. Xu, L.: Formal Verification of Differential Privacy in Concurrent Systems. Ph.D. thesis, Ecole Polytechnique (Palaiseau, France) (2015)
31. Xu, L., Chatzikokolakis, K., Lin, H.: Metrics for differential privacy in concurrent systems. In: FORTE. pp. 199–215. Springer (2014)