Metrics for Differential Privacy in Concurrent Systems

Lili Xu[1,3,4,5], Konstantinos Chatzikokolakis[2,3],
Huimin Lin[5], and Catuscia Palamidessi[1,3]

[1] INRIA [2] CNRS [3] Ecole Polytechnique
[4] Grad. Univ. [5] Inst. of Software, Chinese Acad. of Sci.

Abstract. Originally proposed for privacy protection in the context of statistical databases, differential privacy is now widely adopted in various models of computation. In this paper we investigate techniques for proving differential privacy in the context of concurrent systems. Our motivation stems from the work of Tschantz et al., who proposed a verification method based on proving the existence of a stratified family of bijections between states, that can track the privacy leakage, ensuring that it does not exceed a given leakage budget. We improve this technique by investigating state properties which are more permissive and still imply differential privacy. We consider three pseudometrics on probabilistic automata: The first one is essentially a reformulation of the notion proposed by Tschantz et al. The second one is a more liberal variant, still based on the existence of a family of bijections, but relaxing the relation between them by integrating the notion of amortization, which results into a more parsimonious use of the privacy budget. The third one aims at relaxing the bijection requirement, and is inspired by the Kantorovich-based bisimulation metric proposed by Desharnais et al. We cannot adopt the latter notion directly because it does not imply differential privacy. Thus we propose a multiplicative variant of it, and prove that it is still an extension of weak bisimulation. We show that for all the pseudometrics the level of differential privacy is continuous on the distance between the starting states, which makes them suitable for verification. Moreover we formally compare these three pseudometrics, proving that the latter two metrics are indeed more permissive than the first one, but incomparable with each other, thus constituting two alternative techniques for the verification of differential privacy.

Keywords: differential privacy, probabilistic automata, bisimulation metrics, verification.

1 Introduction

Differential privacy [12] was originally proposed for privacy protection in the context of statistical databases, but nowadays it is becoming increasingly popular in many other fields, ranging from programming languages [21] to social networks [20] and geolocation [19]. One of the reasons of its success is its independence from side knowledge, which makes it robust to attacks based on combining various sources of information.

In the original definition, a query mechanism $\mathcal{A}$ is ϵ-differentially private if for any two databases u_1 and u_2 which differ only for one individual (one raw), and any property Z, the probability distributions of $\mathcal{A}(u_1), \mathcal{A}(u_2)$ differ on Z at most by e^ϵ, namely, $\Pr[\mathcal{A}(u_1) \in Z] \leq e^\epsilon \cdot \Pr[\mathcal{A}(u_2) \in Z]$. This means that the presence (or the

data) of an individual cannot be revealed by querying the database. In [7], the principle of differential privacy has been formally extended to measure the degree of protection of secrets in more general settings.

In this paper we deal with the problem of verifying differential privacy properties for concurrent systems, modeled as probabilistic automata admitting both nondeterministic and probabilistic behavior. In such systems, reasoning about the probabilities requires *solving* the nondeterminism first, and to such purpose the usual technique is to consider functions, called *schedulers*, which select the next step based on the history of the computation. However, in our context, as well as in security in general, we need to restrict the power of the schedulers and make them unable to distinguish between secrets in the histories, or otherwise they would plainly reveal them by their choice of the step. See for instance [6, 15, 8] for a discussion on this issue. Thus we consider a restricted class of schedulers, called *admissible schedulers*, following the definition of [2]. Admissibility is introduced to deal with bisimulation-like notions in security contexts: Two bisimilar processes are typically considered to be indistinguishable, yet an unrestricted scheduler could trivially separate them.

The property of differential privacy requires that the observations generated by two different secret values be probabilistically similar. In standard concurrent systems the notion of similarity is usually formalized as an equivalence, preferably preserved under composition, i.e., a congruence. We mention in particular trace equivalence and bisimulation. The first is often used for its simplicity, but in general is not compositional. The second one is a congruence and it is appealing for its proof technique. Process equivalences have been extensively used to formalize security properties like secrecy [1] and noninterference [13, 22, 23].

In probabilistic systems, we need notions which are robust with respect to small variations in the probabilities, and therefore we usually prefer metric notions over equivalences. In their seminal work, Desharnais et al. [11] proposed a pseudometric based on the Kantorovich metric, which is particularly appealing because it extends weak bisimilarity (captured by the property of having distance 0) and it is based on a natural way of relating probability masses distributed on a metric space. It also satisfies the property that the composition does not increase the distance, which can be considered the metric generalization of the congruence property.

In this paper we focus on metrics suitable for verifying differential privacy. Namely, metrics for which the distance between two processes determines an upper bound on the ratio of the probabilities of the respective observables. We start by considering the framework proposed by Tschantz et al. [25], which was explicitly designed for the purpose of verifying differential privacy. Their verification technique is based on proving the existence of an indexed family of bijections between states. The parameter of the starting states, representing the privacy budget, determines the level of differential privacy of the system, which decreases over time by subtracting the absolute difference of probabilities in each step during mutual simulation. Once the balance reaches zero, processes must behave exactly the same. We reformulate this notion in the form of a pseudometric, showing some novel properties as a distance relation.

The above technique is sound, but limited by the strictness of the relation to be proved, which requires a strong correspondence on states and on their probabilities,

and has a rather rigid budget management. The main goal of this paper is to make the technique more permissive by identifying metrics that are more relaxed and still imply an upper bound on the privacy leakage.

The first improvement we propose is based on a thriftier use of the privacy budget. Inspired by the notion of amortisation used in some quantitative bisimulations [17, 18, 9], we propose a new pseudometric which is more permissive than the former one. The idea is that, when constructing the bijections between states, the differences among the probabilities of related states are kept with their sign, and added with their sign through each step. In this way, successive differences can compensate (amortise) each other, and rather than always being consumed, the privacy budget may also be refurbished.

Although this second metric is inspired by amortised bisimulation, it is not an extension of weak bisimulation, since it still requires bijections between the states, which is in general more restrictive than the bisimilarity relation. It is therefore natural to explore also the use of bisimulation metrics, and to consider the metric à la Kantorovic proposed in [11], which represents a cornerstone in this area. However, we cannot use directly the metric of [11] because it does not imply differential privacy: the problem is that the difference in probabilities in this metric is accounted for additively, while differential privacy is a property about their ratio. Thus, we propose a multiplicative variant of it, and obtain a pseudometric that, to the best of our knowledge, is new.

We show that the distance in this pseudometric can be computed using linear programming solution method (by simply using its dual form). Intuitively, in the context of *transportation problem*, Kantorovich metric gives the lowest total cost of transporting the mass of one distribution μ to the other distribution μ', while our variant is used to achieve the lowest cost of transportation per unit mass of μ', namely the optimal efficiency. We also show that our variant satisfies most of the properties of the metric in [11]: in particular, it can be characterized by a fixed-point construction, and it extends weak bisimilarity.

While this third metric is more liberal than the first one, it is not comparable with the second. The reason is that the management of the budget in the second metric follows the spirit trace semantics (which is coarser than bisimulation), in that the budget gets amortised by adding positive or negative differences trough the step-by-step comparison of two traces.

More related Work. Verification of differential privacy has become an active area of research. Among the approaches based on formal methods, we mention those based on type-systems [21, 14] and logical formulations [4, 3]

In a previous paper [26], one of the authors has developed a compositional method for proving differential privacy in a probabilistic process calculus. The technique there is rather different from the ones presented in paper: the idea is based on decomposing a process in simpler processes, computing the level of privacy of these, and combining them to obtain the level of privacy of the original program.

Contribution. The main contributions of this paper can be summarized as follows:

- We reformulate the notion of approximate similarity proposed in [25] in terms of a pseudometric and we study the properties of the distance relation.

- We propose the second pseudometric which is more liberal than the former one, in the sense that the total differences of probabilities get amortised during the mutual simulation.
- We propose the third pseudometric, which a multiplicative variant of Kantorovich-based bisimulation metric in [11].
- We show that for all the pseudometrics the level of differential privacy is continuous on the distance between the starting states, which makes them suitable for verification.
- We compare these three pseudometrics and study their relations with weak bisimilarity, proving that the latter two metrics are more permissive than the first one, but incomparable with each other.

Plan of the Paper. In the next section we recall some preliminary notions about probabilistic automata, differential privacy and pseudometrics. Sections 3, 4, 5 introduce respectively the first, second and third pseudometrics, and prove for differential privacy the soundness of the verification technique with respect to each of these three pseudometrics. In Section 6 we compare these three metrics and study their relations with weak bisimilarity. Section 7 concludes. Proofs can be found in the appendix.

2 Preliminaries

2.1 Probabilistic automata

Given a set X, we denote by $Disc(X)$ the set of discrete sub-probability measures over X; the support of a measure μ is defined as $supp(\mu) = \{x \in X | \mu(x) > 0\}$. A *probabilistic automaton* (henceforth PA) $\mathcal{A}$ is a tuple $(S, \overline{s}, A, D)$ where S is a finite set of *states*, $\overline{s} \in S$ is the *start state*, A is a finite set of action *labels*, and $D \subseteq S \times A \times Disc(S)$ is a *weak transition relation*. We write $s \stackrel{a}{\Longrightarrow} \mu$ for $(s, a, \mu) \in D$, and we denote by $act(d)$ the action of the transition $d \in D$. Note that $\Longrightarrow$ is typically obtained from an original transition relation by merging τ transitions (see, for instance, [11]). A PA $\mathcal{A}$ is *fully probabilistic* if from each state of $\mathcal{A}$ there is at most one transition available.

A *(weak) execution* α of a PA is a (possibly infinite) sequence $s_0 a_1 s_1 a_2 s_2 \ldots$ of alternating states and labels, such that for each $i : s_i \stackrel{a_{i+1}}{\Longrightarrow} \mu_{i+1}$ and $\mu_{i+1}(s_{i+1}) > 0$. We use $lstate(\alpha)$ to denote the last state of a finite execution α. We use $Exec^*(\mathcal{A})$ and $Exec(\mathcal{A})$ to represent the set of finite weak executions and of all weak executions of $\mathcal{A}$, respectively. A *scheduler* of a PA $\mathcal{A} = (S, \overline{s}, A, D)$ is a function $\zeta : Exec^*(\mathcal{A}) \mapsto D$ such that $\zeta(\alpha) = s \stackrel{a}{\Longrightarrow} \mu \in D$ implies that $s = lstate(\alpha)$. The idea is that a scheduler selects a transition among the ones available in D, basing its decision on the history of the execution. The *(weak) execution tree* of $\mathcal{A}$ relative to the scheduler ζ, denoted by $\mathcal{A}_\zeta$, is a fully probabilistic automaton $(S', \overline{s}', A', D')$ such that $S' \subseteq Exec^*(\mathcal{A})$, $\overline{s}' = \overline{s}$, $A' = A$, and $\alpha \stackrel{a}{\Longrightarrow} \nu \in D'$ if and only if $\zeta(\alpha) = lstate(\alpha) \stackrel{a}{\Longrightarrow} \mu$ for some μ and $\nu(\alpha a s) = \mu(s)$. Intuitively, $\mathcal{A}_\zeta$ is produced by unfolding the executions of $\mathcal{A}$ and resolving all non-deterministic choices using ζ. Note that $\mathcal{A}_\zeta$ is a simple and fully probabilistic automaton. We use α with primes and indices to range over states in an execution tree.

A *(weak) trace* is a sequence of labels in $A^* \cup A^\omega$ obtained from executions by removing the states. We use $[\,]$ to represent the empty trace, and $^\frown$ to concatenate two traces. A state α of $\mathcal{A}_\zeta$ induces a probability measure over traces as follows. The basic measurable events are the cones of finite traces, where the cone of a finite trace $\boldsymbol{t}$, denoted by $C_{\boldsymbol{t}}$, is the set $\{\boldsymbol{t}' \in A^* \cup A^\omega | \boldsymbol{t} \leq \boldsymbol{t}'\}$, where $\leq$ is the standard prefix preorder on sequences. The probability of a cone $C_{\boldsymbol{t}}$ induced by state α, denoted by $\Pr_\zeta[\alpha \rhd \boldsymbol{t}]$, is defined recursively as follows.

$$\Pr_\zeta[\alpha \rhd \boldsymbol{t}] = \begin{cases} 1 & \text{if } \boldsymbol{t} = [\,], \\ 0 & \text{if } \boldsymbol{t} = a^\frown \boldsymbol{t}' \text{ and } act(\zeta(\alpha)) \neq a, \\ \sum_{s_i \in supp(\mu)} \mu(s_i) \Pr_\zeta[\alpha a s_i \rhd \boldsymbol{t}'] & \\ & \text{if } \boldsymbol{t} = a^\frown \boldsymbol{t}' \text{ and } \zeta(\alpha) = s \stackrel{a}{\Longrightarrow} \mu. \end{cases} \tag{1}$$

Admissible schedulers. In concurrent systems containing both non-deterministic and probabilistic behavior, it is well-known that the scheduler (i.e. the entity resolving the non-determinism) can easily break many security and privacy properties by choosing different transition based on a secret value. As a consequence, to perform a meaningful analysis one needs to restrict to a class of *admissible* schedulers, which do not exhibit such a behavior. Thus we consider a restricted class of schedulers, called *admissible schedulers*, following the definition of [2]. Essentially this definition requires that whenever given two *adjacent* states s, s', namely, differing only for the choice for some secret value, then the choice made by the scheduler on s and s' should be consistent, i.e. the scheduler should not be able to make a different choice on the basis of the secret. Note that in [25] scheduling is not an issue since non-determinism is not allowed.

Pseudometrics on states. A pseudometric[1] on S is a function $m : S^2 \to \mathbb{R}$ satisfying the following properties: $m(s,s) = 0$ (reflexivity), $m(s,t) = m(t,s)$ (symmetry) and $m(s,t) \leq m(s,u) + m(u,t)$ (triangle inequality). Let $\mathcal{M}$ denote the set of all pseudometrics on S, with the ordering $m_1 \preceq m_2$ iff $\forall s,t : m_1(s,t) \geq m_2(s,t)$ (note that the order is reversed). It can be shown that $(\mathcal{M}, \preceq)$ is a complete lattice.

2.2 Differential privacy

Differential privacy [12] was originally defined in the context of statistical databases, by requiring that a mechanism (i.e. a probabilistic query) gives similar answers on *adjacent* databases, that is those differing on a single row. More precisely, a mechanism $\mathcal{K}$ satisfies *ϵ-differential privacy* iff for all adjacent databases x, x': $\Pr[\mathcal{K}(x) \in Z] \leq e^\epsilon \cdot \Pr[\mathcal{K}(x') \in Z]$ for all $Z \subseteq range(\mathcal{K})$.

In this paper, we study concurrent systems taking a secret as input and producing an observable trace as output. Let U be a set of secrets and $\sim$ an adjacency relation on U, where $u \sim u'$ denotes the fact that two close secrets u, u' should not be easily distinguished by the adversary after seeing observable traces. A *concurrent system* $\mathcal{A}$ is a mapping of secrets to probabilistic automata, where $\mathcal{A}(u), u \in U$ is the automaton

[1] Unlike a metric, points in a pseudometric need not be distinguishable; that is, one may have $m(s,t) = 0$ for distinct values $s \neq t$.

modelling the behaviour of the system when running on u. Differential privacy can be directly adapted to this context:

Definition 1 (Differential Privacy). *A concurrent system $\mathcal{A}$ satisfies* ϵ-differential privacy *(DP) iff for any $u \sim u'$, any finite trace $\boldsymbol{t}$ and any admissible scheduler ζ:*

$$\Pr_\zeta[\mathcal{A}(u) \rhd \boldsymbol{t}] \leq e^\epsilon \cdot \Pr_\zeta[\mathcal{A}(u') \rhd \boldsymbol{t}]$$

3 The accumulative bijection pseudometric

In this section, we present the first pseudometric based on a reformulation of the relation family proposed in [25]. We reformulate their notion in the form of an approximate bisimulation relation, named *accumulative bisimulation*, and the use it to construct a pseudometric on the state space.

We start by defining an approximate lifting operation that lifts a relation over states to a relation over distributions. We use D to simply differentiate notions of this section from the following sections. Intuitively, we use a parameter ϵ to represent the total privacy leakage budget. A parameter c ranging over $[0, \epsilon]$, starting from 0, records the current amount of leakage and increasing over time by adding the maximum absolute difference of probabilities, denoted by σ, in each step during mutual simulation. Once c reaches the budget bound ϵ, processes must behave exactly the same. Since the total bound is ϵ, only a total of ϵ privacy can be leaked, a fact that will be used later to verify differential privacy.

Definition 2. *Let $\epsilon \geqslant 0$, $c \in [0, \epsilon]$, $\mathcal{R} \subseteq S \times S \times [0, \epsilon]$. The* D-approximate lifting *of $\mathcal{R}$ up to c, denoted by $\mathcal{L}^D(\mathcal{R}, c)$, is the relation on $Disc(S)$ defined as:*

$$\mu \mathcal{L}^D(\mathcal{R}, c) \mu' \quad \textit{iff} \quad \exists \textit{ bijection } \beta : supp(\mu) \to supp(\mu') \textit{ such that}$$
$$\forall s \in supp(\mu) : (s, \beta(s), c + \sigma) \in \mathcal{R} \quad \textit{where} \quad \sigma = \max_{s \in supp(\mu)} |\ln \frac{\mu(s)}{\mu'(\beta(s))}|$$

This lifting allows us to define an approximate bisimulation relation:

Definition 3 (Accumulative bisimulation). *A relation $\mathcal{R} \subseteq S \times S \times [0, \epsilon]$ is an* ϵ-accumulative bisimulation *iff for all $(s, t, c) \in \mathcal{R}$:*

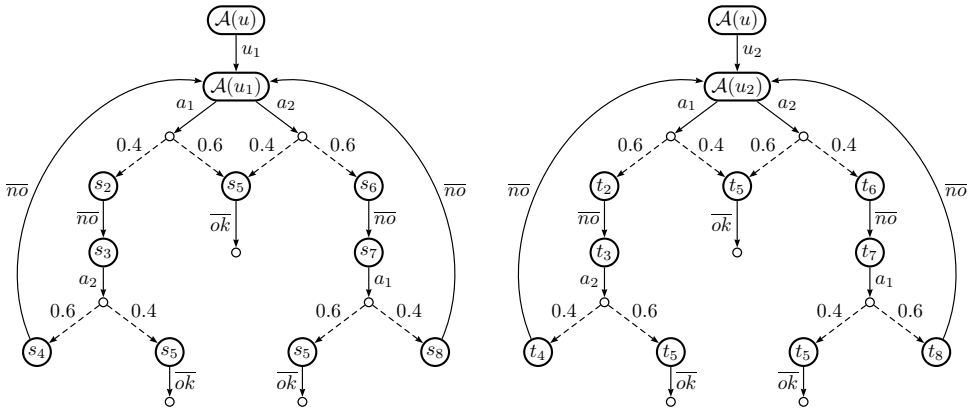
1. *$s \stackrel{a}{\Longrightarrow} \mu$ implies $t \stackrel{a}{\Longrightarrow} \mu'$ with $\mu \mathcal{L}^D(\mathcal{R}, c) \mu'$*
2. *$t \stackrel{a}{\Longrightarrow} \mu'$ implies $s \stackrel{a}{\Longrightarrow} \mu$ with $\mu \mathcal{L}^D(\mathcal{R}, c) \mu'$*

We can now define a pseudometric based on accumulative bisimulation as:

$$m^D(s, t) = \min\{\epsilon \,|\, (s, t, 0) \in \mathcal{R} \text{ for some } \epsilon\text{-accumulative bisimulation } \mathcal{R}\}$$

Proposition 1. *m^D is a pseudometric, that is:*

1. *(reflexivity) $m^D(s, s) = 0$*
2. *(symmetry) $m^D(s_1, s_2) = m^D(s_2, s_1)$*
3. *(triangle inequality) $m^D(s_1, s_3) \leq m^D(s_1, s_2) + m^D(s_2, s_3)$*



(a) $\mathcal{A}(u_1)$ (b) $\mathcal{A}(u_2)$

Fig. 1: Distance between $\mathcal{A}(u_1)$ and $\mathcal{A}(u_2)$, m^D gives ∞, while m^A gives $\ln \frac{9}{4}$.

Verification of differential privacy using m^D. As already shown in [25], the closeness of processes in the relation family implies a level of differential privacy. We here restate this result in terms of the metric m^D.

Lemma 1. *Given a PA $\mathcal{A}$, let $\mathcal{R}$ be an ϵ-accumulative bisimulation, $c \in [0, \epsilon]$, let ζ be an admissible scheduler, $\boldsymbol{t}$ be a finite trace, α_1, α_2 two finite executions of $\mathcal{A}$. If $(lstate(\alpha_1), lstate(\alpha_2), c) \in \mathcal{R}$, then*

$$\frac{1}{e^{\epsilon - c}} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^{\epsilon - c}$$

The above lemma shows that in an ϵ-accumulative bisimulation, two states related by a current leakage amount c, produce distributions over the same trace that only deviate by a factor $(\epsilon - c)$ representing the remaining amount of leakage. Then it is easy to get that the level of differential privacy is continuous on m^D.

Theorem 1. *A concurrent system $\mathcal{A}$ is ϵ-differentially private if $m^D(\mathcal{A}(u), \mathcal{A}(u')) \leq \epsilon$ for all $u \sim u'$.*

4 The amortized bijection pseudometric

As shown in the previous section, m^D is useful for verifying differential privacy. However, a drawback of this metric is that the definition of accumulative bisimulation is too restrictive: first, the amount of leakage is only accumulated, independently from whether the difference in probabilities is negative or positive. Moreover, the accumulation is same for all branches, and equal to the worst branch, although the actual difference on some branch might be small. As a consequence, m^D is inapplicable in several systems, as shown by the following example.

Example 1. Consider a concurrent system $\mathcal{A}$ shown in Fig. 1. Consider an admissible scheduler always choosing for $\mathcal{A}(u_1)$ the a_1-branch (the case for the a_2-branch is similar), thus scheduling for $\mathcal{A}(u_2)$ also the a_1-branch. It is easy to see that the ratio of probabilities for $\mathcal{A}(u_1)$ and $\mathcal{A}(u_2)$ producing the same finite sequences $(a_1\overline{no}\,a_2\,\overline{no})^*$ is $(\frac{0.4\times0.6}{0.6\times0.4})^* = 1$. For the rest sequences $(a_1\overline{no}\,a_2\,\overline{no})^*a_1\overline{ok}$ and $(a_1\overline{no}\,a_2\,\overline{no})^*a_1\overline{no}\,a_2\,\overline{ok}$, we can check that the ratios are bounded by $\frac{9}{4}$. Thus, $\mathcal{A}$ satisfies $\ln\frac{9}{4}$-differential privacy. However, we can not find an accumulative bisimulation with a bounded ϵ between $\mathcal{A}(u_1)$ and $\mathcal{A}(u_2)$. The problem lies in that the leakage amount is always accumulated by adding the absolute differences during cyclic simulations, resulting in a convergence to ∞.

In order to obtain a more relaxed metric, we employ the *amortised bisimulation* relation of [17, 18]. The main intuition behind this notion is that the privacy leakage budget in each simulation step may be either reduced due to a negative difference of probabilities, or increase due to a positive difference. Hence, the long-term budget gets amortised, in contrast to the accumulative bisimulation in which the budget is always consumed. We start by defining the corresponding lifting, using A to represent amortised bisimulation-based notions. Note that the current leakage c ranges over $[-\epsilon, \epsilon]$.

Definition 4. *Let $\epsilon \geqslant 0$, $c \in [-\epsilon, \epsilon]$, $\mathcal{R} \subseteq S \times S \times [-\epsilon, \epsilon]$. The* A-approximate lifting *of $\mathcal{R}$ up to c, denoted by $\mathcal{L}^A(\mathcal{R}, c)$, is a relation on $Disc(S)$ defined as:*

$$\mu\mathcal{L}^A(\mathcal{R}, c)\mu' \quad \textit{iff} \quad \exists \textit{ bijection } \beta : supp(\mu) \to supp(\mu') \textit{ such that}$$
$$\forall s \in supp(\mu) : (s, \beta(s), c + \ln\frac{\mu(s)}{\mu'(\beta(s))}) \in \mathcal{R}$$

Note that if $\ln\frac{\mu(s)}{\mu'(\beta(s))}$ is positive, then after this mutual step, the current leakage for s and $\beta(s)$ gets increased, otherwise decreased. We are now ready to define amortised bisimulation.

Definition 5 (Amortised bisimulation). *A relation $\mathcal{R} \subseteq S \times S \times [-\epsilon, \epsilon]$ is an* ϵ-amortised bisimulation *iff for all $(s, t, c) \in \mathcal{R}$:*

1. *$s \stackrel{a}{\Longrightarrow} \mu$ implies $t \stackrel{a}{\Longrightarrow} \mu'$ with $\mu\mathcal{L}^A(\mathcal{R}, c)\mu'$*
2. *$t \stackrel{a}{\Longrightarrow} \mu'$ implies $s \stackrel{a}{\Longrightarrow} \mu$ with $\mu\mathcal{L}^A(\mathcal{R}, c)\mu'$*

Note that the coalgebraic character of amortised bisimulation notion has been justified in [9], ensuring that it inherits most of the good properties of quantitative bisimulation semantics, such as the existence of a fixed-point characterization.

Similarly to the previous section, we can finally define a pseudometric on states as:

$$m^A(s, t) = \min\{\epsilon \mid (s, t, 0) \in \mathcal{R} \text{ for some } \epsilon\text{-amortized bisimulation } \mathcal{R}\}$$

Proposition 2. *m^A is a pseudometric.*

Verification of differential privacy using m^A. We now show that m^A can be used to verify differential privacy.

Lemma 2. *Given a PA $\mathcal{A}$, let $\mathcal{R}$ be an ϵ-amortised bisimulation, $c \in [-\epsilon, \epsilon]$, let ζ be an admissible scheduler, $\boldsymbol{t}$ be a finite trace, α_1, α_2 two finite executions of $\mathcal{A}$. If $(lstate(\alpha_1), lstate(\alpha_2), c) \in \mathcal{R}$, then*

$$\frac{1}{e^{\epsilon+c}} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^{\epsilon-c}$$

Note that there is a subtle difference between Lemmas 1 and 2, in that the left-hand bound is $e^{\epsilon+c}$ instead of $e^{\epsilon-c}$. This comes from the amortised nature of $\mathcal{R}$. We can now show that differential privacy is continuous on m^A as well.

Theorem 2. *A concurrent system $\mathcal{A}$ is ϵ-differentially private if $m^A(\mathcal{A}(u), \mathcal{A}(u')) \leq \epsilon$ for all $u \sim u'$.*

Example 2 (Example 1 revisited). Consider again the concurrent system shown in Fig. 1. Let S and T denote the state space of $\mathcal{A}(u_1)$ and $\mathcal{A}(u_2)$, respectively. Let $\mathcal{R} \subseteq S \times T \times [\ln \frac{4}{9}, \ln \frac{9}{4}]$. It is straightforward to check according to Def. 5 that the following relation is an amortised bisimulation between $\mathcal{A}(u_1)$ and $\mathcal{A}(u_2)$.

$$\begin{aligned}\mathcal{R} = \{\ & (\mathcal{A}(u_1), \mathcal{A}(u_2), 0),\\ & (s_2, t_2, \ln \tfrac{2}{3}), \qquad (s_6, t_6, \ln \tfrac{3}{2}),\\ & (s_5, t_5, \ln \tfrac{3}{2}), \qquad (s_5, t_5, \ln \tfrac{2}{3}),\\ & (s_3, t_3, \ln \tfrac{2}{3}), \qquad (s_7, t_7, \ln \tfrac{3}{2}),\\ & (s_4, t_4, 0), \qquad (s_8, t_8, 0),\\ & (s_5, t_5, \ln \tfrac{4}{9}), \qquad (s_5, t_5, \ln \tfrac{9}{4})\ \}\end{aligned}$$

Thus $m^A(\mathcal{A}(u_1), \mathcal{A}(u_2)) = \ln \frac{9}{4}$, by Theorem 2, $\mathcal{A}$ is $\ln \frac{9}{4}$-differentially private.

5 The multiplicative variant of the Kantorovich pseudometric

The Kantorovich metric (shown in the left column of Fig. 2) is a widely used construction for lifting a metric from a set to distributions over this set. In a well-known work, Desharnais et al. [11] use this lifting and a fixpoint construction to define a metric on states, denoted by m^O, which characterizes weak bisimulation.

Since our goal is to use metrics for verifying differential privacy, a natural question that arises is whether m^O can be employed for our goal. However, when trying to use m^O for this purpose, one quickly realizes that the "additive" nature of the Kantorovich metric makes it inadequate for verifying a "multiplicative" property such as differential privacy. An example of this behaviour is given later in this section.

As a consequence, we propose a *multiplicative* variant of the Kantorovich metric, which gives rise to a third pseudometric m^K. On the one hand, m^K inherits most of the appealing behaviors of m^O, while being adequate for verifying differential privacy.

	Kantorovich metric	The multiplicative variant
Primal	maximize $\sum_i(\mu(s_i) - \mu'(s_i))x_i$ subject to $\quad \forall i.\ 0 \leq x_i \leq 1$ $\forall i, j.\ x_i - x_j \leq m(s_i, s_j)$	maximize $\lvert \ln \frac{\sum_i \mu(s_i)x_i}{\sum_i \mu'(s_i)x_i} \rvert$ subject to $\quad \forall i.\ 0 \leq x_i \leq 1$ $\forall i, j.\ x_i \leq e^{m(s_i,s_j)}x_j$
Dual	minimize $\sum_{i,j} l_{ij}m(s_i, s_j) + \sum_i x_i + \sum_j y_j$ subject to $\quad \forall i.\ \sum_j l_{ij} + x_i = \mu(s_i)$ $\forall j.\ \sum_i l_{ij} + y_j = \mu'(s_j)$ $\forall i, j.\ l_{ij}, x_i, y_j \geq 0$	minimize $\ln z$ subject to $\quad \forall i. \sum_j l_{ij} - r_i = \mu(s_i)$ $\forall j.\ \sum_i l_{ij}e^{m(s_i,s_j)} - r_j \leq z \cdot \mu'(s_j)$ $\forall i, j.\ l_{ij}, r_i \geq 0$

Fig. 2: The Kantorovich metric and its multiplicative variant.

5.1 Adapting the Kantorovich metric

We now give a multiplicative variant of the Kantorovich metric, which computes the distance between probability distributions in a multiplicative sense, namely, with respect to the ratio between the distributions. With a slight abuse use of notation, we use m to denote both the original metric on states and its lifting.

Definition 6. *Let $m \in \mathcal{M}$. Let μ, μ' be distributions on states. The metric $m(\mu, \mu')$ is given by the solution of the primal optimization program shown in Fig. 2 (multiplicative variant).*

The primal program can be converted (the details are given in the appendix) to a quasi-linear dual program (Fig. 2), (For the sake of simplicity, Fig. 2 only shows the dual program of $\ln \sum_i \mu(s_i)x_i - \ln \sum_i \mu'(s_i)x_i$. The dual program for $\ln \sum_i \mu'(s_i)x_i - \ln \sum_i \mu(s_i)x_i$ can be obtained by simply switching the roles of μ and μ'.) which can be solved using any linear programming method. In addition, this dual form will be essential to prove the correspondence with weak bisimilarity (in Section 6.1) and give an intuitive interpretation latter.

Intuitive difference between the Kantorovich metric and our variant. First let us recall the intuition of the Kantorovich metric. Consider its dual program in Fig. 2. The Kantorovich metric is usually interpreted as a *transportation problem*. Intuitively, l_{ij} can be understood as a transportation of l_{ij} unit mass from a location $s_i \in supp(\mu)$ to a location $s_j \in supp(\mu')$, the cost of moving one unit of mass from s_i to s_j is represented by (a function of) the distance $m(s_i, s_j)$. Then the Kantorovich distance gives the lowest total cost of transporting the mass of μ to μ'.

Similarly, the dual program of the multiplicative variant can be used to obtain an intuitive interpretation. By simple transformation, we obtain $z \geq (\sum_{i,j} l_{ij}e^{m(s_i,s_j)} - \sum_i r_i)/\sum_j \mu'(s_j)$, where l_{ij} and $m(s_i, s_j)$ are read in the same way as above. Now we can see that our variant is used to achieve the lowest cost of transportation per unit mass of μ', the ratio representing the optimal efficiency.

We can finally show that the lifted metric is indeed a metric.

Lemma 3. *Let $m \in \mathcal{M}$. The Kantorovich lifting $m(\mu, \mu')$ is a metric on $Disc(S)$. Moreover, $m \preceq m'$ implies $m(\mu, \mu') \geq m'(\mu, \mu')$.*

5.2 The m^K metric on states

We are now ready to use the Kantorovich lifting to define a third pseudometric m^K on states. We start by defining the concept of a K-state-metric:

Definition 7. *$m \in \mathcal{M}$ is a* K-state-metric *if, for any ϵ, $m(s,t) \leq \epsilon$ implies that if $s \stackrel{a}{\Longrightarrow} \mu$ then there exists some μ' such that $t \stackrel{a}{\Longrightarrow} \mu'$ and $m(\mu, \mu') \leq \epsilon$.*

Note that in the above definition, the "vice-versa" case is covered by the fact that m is also a metric. By $m(s,t) \leq \epsilon$ we have $m(t,s) \leq \epsilon$, which implies that if $t \stackrel{a}{\Longrightarrow} \mu'$ then there exists some μ such that $s \stackrel{a}{\Longrightarrow} \mu$ and $m(\mu', \mu) \leq \epsilon$.

Then m^K is defined as the greatest K-state-metric:

$$m^K = \bigsqcup \{m \in \mathcal{M} \mid m \text{ is a } K\text{-state-metric}\}.$$

A fixed-point characterization. Next, we characterize m^K as the greatest fixed-point of a monotone function on a complete lattice. This approach was first proposed by Desharnais et al. [11] for labelled concurrent Markov chains.

Definition 8. *Define F, a functional on $\mathcal{M}$ as follows. $F(m)(s,t) \leq \epsilon$ if and only if:*

1. $(\forall s \stackrel{a}{\Longrightarrow} \mu)(\exists t \stackrel{a}{\Longrightarrow} \mu')[m(\mu, \mu') \leq \epsilon]$.
2. $(\forall t \stackrel{a}{\Longrightarrow} \mu')(\exists s \stackrel{a}{\Longrightarrow} \mu)[m(\mu, \mu') \leq \epsilon]$.

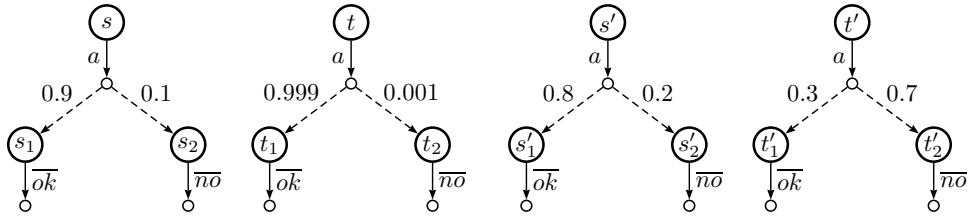
The triangle inequality on $F(m)$ follows from the triangle inequality on m extended to distributions (see Lemma 3). By adapting the proofs of the analogous results in [11, 10], $F(m)$ is well-defined. By directly checking the definition of F, it is easy to see that m is a K-state-metric if and only if $m \preceq F(m)$. Henceforth, we have that m^K is exactly the greatest pre-fixed-point of F, namely, $m^K = \bigsqcup \{m \in \mathcal{M} | m \preceq F(m)\}$. By Lemma 3, the ordering on pseudometrics is preserved when metrics are lifted from states to distributions over states, it is routine to get that F is monotone on $\mathcal{M}$. Therefore, since $(\mathcal{M}, \preceq)$ is a complete lattice, we can apply Tarski's fixed point theorem [24], which ensures that F has a maximum fixed-point. In addition, the finite-stateness of the PA $\mathcal{A}$ ensures that closure ordinal of F is ω (cf: [11], Lemma 3.10). Hence we can proceed in a standard way to show that m^K is indeed the largest fixed-point of F, and is given by $m^K = \sqcap_i m_i$, where $m_0 = \top$ and $m_{i+1} = F(m_i)$.

5.3 Verification of differential privacy using m^K

We then show that m^K is suitable for verifying differential privacy.

Lemma 4. *Given a PA $\mathcal{A}$, let ζ be an admissible scheduler, $\boldsymbol{t}$ be a finite trace, and α_1, α_2 be two finite executions in $\mathcal{A}$. If $m^K(lstate(\alpha_1), lstate(\alpha_2)) \leq \epsilon$, then:*

$$\frac{1}{e^\epsilon} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^\epsilon$$



(a) $m^O(s,t) = 0.099$ while $\epsilon = \ln 100$. (b) $m^O(s',t') = 0.5$ while $\epsilon' = \ln 3.5$.

Fig. 3: The pseudometric of [11] does not imply differential privacy

Finally, we can show that m^K can be used to verify differential privacy.

Theorem 3. *A concurrent system $\mathcal{A}$ is ϵ-differentially private if $m^K(\mathcal{A}(u), \mathcal{A}(u')) \leq \epsilon$ for all $u \sim u'$.*

Differential privacy and the standard Kantorovich metric. We now revisit the question of using the original Kantorovich metric m^O for verifying privacy. Recall that m^O, introduced in [11], is defined in the same way as m^K, but using the standard Kantorovich lifting. The example below shows that m^O may be very different from the level of differential privacy.

Example 3. Consider two processes s, t shown in Fig. 3 (a), compute $m^O(s,t) = 0.1 - 0.001 = 0.099$ while the level of differential privacy $\epsilon = \ln \frac{0.1}{0.001} = \ln 100$. Consider another two processes s', t' shown in Fig. 3 (b), compute $m^O(s',t') = 0.7 - 0.2 = 0.5$ while the level of differential privacy $\epsilon' = \ln \frac{0.7}{0.2} = \ln 3.5$. Using the original Kantorovich metric, s and t are considered more indistinguishable than s' and t', in sharp contrast to the corresponding differential privacy levels.

This behaviour is due to the additive nature of m^O. In fact, we can show that m^O induces a bound on the difference between the probabilities of producing a trace.

Lemma 5. *Given a PA $\mathcal{A}$, let ζ be an admissible scheduler, let $\boldsymbol{t}$ be a finite trace, and α_1, α_2 be two finite executions in $\mathcal{A}$. If $m^O(lstate(\alpha_1), lstate(\alpha_2)) \leq \epsilon$, then:*

$$|\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] - \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]| \leq \epsilon$$

6 Comparing the three metrics

In this section, we compare the three metrics, showing that the latter two are indeed more liberal than the first one, although incomparable to each other. Moreover,we investigate their relation with weak bisimilarity.

We first show that m^A is bounded by m^D.

Lemma 6. $m^D \preceq m^A$.

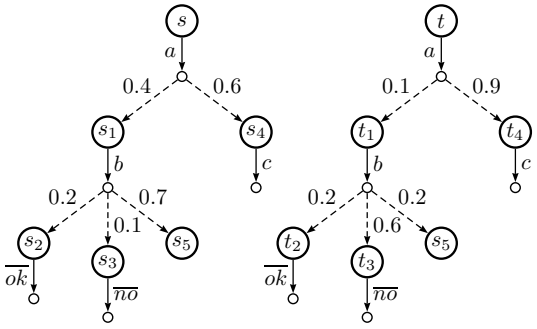


Fig. 4: $m^K(s,t) > m^A(s,t)$

Note the converse does not hold, since Examples 1 and 2 already show cases in which m^D is infinite while m^A is finite.

Then, we show that m^K is also bounded by m^D.

Lemma 7. $m^D \preceq m^K$.

To show that m^A, m^K are incomparable to each other, we first show a toy example in which $m^K(s,t) > m^A(s,t)$.

Example 4. Consider two processes s, t shown in Fig. 4. Through computations (see Appendix), we obtain that $m^K(s,t) = \ln 24$, while $m^A(s,t) = \ln 14$ which is finer. The idea behind this example is that m^A works well in the scenario where by amortising the total differences of probabilities, the resulting ratio gets smaller than by accumulating.

For the converse, note that from the definitions of m^A, a strong bijection relation is required in each mutual simulation. We can easily find counterexamples that do not have this bijection relation between the distributions of the starting states, making m^A infinite. On the other hand, m^K is completely permissive on this point.

6.1 Relations with weak bisimilarity

Finally, we show that, similarly to the original Kantorovich metric, the 0 distance of m^K characterizes weak bisimilarity, thus satisfying the criterion on metrics for probabilistic processes of Desharnais et al. [11]. On the other hand, m^D and m^A only imply weak bisimilarity, while the converse direction does not hold because of the strong requirement of the bijections. We adopt the notion of weak bisimilarity proposed in [11], the details can be found in the appendix.

The maximal fixed point of F introduced in Section 5.2 as an alternative characterization of m^K, corresponds to weak bisimilarity. The forward implication of the proof is proved by defining a metric $m(s,t) = 0$ if s and t are weak bisimilar, and ∞ otherwise, and showing that $m \preceq F(m)$. The converse implication proceeds by defining an equivalence relation induced by the metric m^K, i.e. $(s,t) \in \mathcal{R}$ iff $m^K(s,t) = 0$, and showing that $\mathcal{R}$ is a weak bisimulation.

Proposition 3. *The following hold:*

- $m^K(s,t) = 0 \Leftrightarrow s \approx t$
- $m^D(s,t) = 0 \Rightarrow s \approx t$
- $m^A(s,t) = 0 \Rightarrow s \approx t$

7 Conclusion and future work

We have investigated three pseudometrics on states: the first one is a reformulation of the notion proposed in [25], the second one is designed in the sense that the total privacy leakage bound gets amortised, the last metric is built on a multiplicative variant inheriting the merits of Kantorovich metric. Each of the three pseudometrics establish a framework for the formal verification of differential privacy for concurrent systems. Namely, the closer processes are in the metrics, the higher level of differential privacy they can preserve.

We have formally compared these pseudometrics, showing that the latter two are more liberal than the first one. They make improvements on the first metric in two orthogonal ways: one by considering a more careful use of the privacy leakage budget; the other by using more relaxed relation between states, inspired by bisimilarity and by the Kantorovich metric. Through some counterexamples, the second and the third metrics are shown to be incomparable with each other, henceforth they can be considered as two alternative techniques for verification of differential privacy.

In this paper we have mainly focus on developing a basic framework for the formal verification of differential privacy for concurrent systems. In the future we plan to develop more realistic case-studies and applications. We also plan to investigate whether and how we can define a new pseudometric that unifies the merits of the amortised pseudometric and the multiplicative variant of the Kantorovich metric. Finally, we want to investigate whether the distance in our multiplicative variant of the Kantorovich metric is compositional with respect to the typical process algebra operators, in the sense that the distance between two compound terms can be calculated as a function of the distance of the components. This would allow to combine the techniques presented in this paper with the compositional method proposed in [26].

References

1. Martín Abadi and Andrew D. Gordon. A calculus for cryptographic protocols: The spi calculus. *Inf. and Comp.*, 148(1):1–70, 1999.
2. Miguel E. Andrés, Catuscia Palamidessi, Ana Sokolova, and Peter Van Rossum. Information Hiding in Probabilistic Concurrent Systems. *TCS*, 412(28):3072–3089, 2011.
3. Gilles Barthe, George Danezis, Benjamin Grégoire, César Kunz, and Santiago Zanella Béguelin. Verified computational differential privacy with applications to smart metering. In *CSF*, pages 287–301, 2013.
4. Gilles Barthe, Boris Köpf, Federico Olmedo, and Santiago Z. Béguelin. Probabilistic relational reasoning for differential privacy. In *Proc. of POPL*. ACM, 2012.
5. Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic transition systems. In *Proc. of CONCUR'01*, pages 336–350, London, UK, 2001. Springer.

6. Ran Canetti, Ling Cheung, Dilsun Kaynar, Moses Liskov, Nancy Lynch, Olivier Pereira, and Roberto Segala. Task-structured probabilistic i/o automata. In *Proc. of WODES*, 2006.
7. Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. Broadening the scope of differential privacy using metrics. In *Privacy Enhancing Technologies*, pages 82–102, 2013.
8. Konstantinos Chatzikokolakis and Catuscia Palamidessi. Making random choices invisible to the scheduler. In *Proc. of CONCUR*, volume 4703 of *LNCS*, pages 42–58. Springer, 2007.
9. David de Frutos-Escrig, Fernando Rosa-Velardo, and Carlos Gregorio-Rodríguez. New bisimulation semantics for distributed systems. In *FORTE*, pages 143–159, 2007.
10. Yuxin Deng, Tom Chothia, Catuscia Palamidessi, and Jun Pang. Metrics for action-labelled quantitative transition systems. In *Proc. of QAPL*, volume 153 of *ENTCS*, pages 79–96. Elsevier, 2006.
11. Josee Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proc. of LICS*, pages 413–422. IEEE, 2002.
12. Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming, 33rd Int. Colloquium, ICALP 2006, Proceedings, Part II*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.
13. Riccardo Focardi and Roberto Gorrieri. Classification of security properties (part i: Information flow). In *FOSAD*, pages 331–396, 2000.
14. Marco Gaboardi, Andreas Haeberlen, Justin Hsu, Arjun Narayan, and Benjamin C. Pierce. Linear dependent types for differential privacy. In *POPL*, pages 357–370, 2013.
15. Flavio D. Garcia, Peter van Rossum, and Ana Sokolova. Probabilistic anonymity and admissible schedulers, 2007. arXiv:0706.1019v1.
16. Joseph A. Goguen and José Meseguer. Security policies and security models. In *IEEE Symposium on Security and Privacy*, pages 11–20, 1982.
17. Astrid Kiehn and S. Arun-Kumar. Amortised bisimulations. In *FORTE*, pages 320–334, 2005.
18. Gerald Lüttgen and Walter Vogler. Bisimulation on speed: A unified approach. *Theor. Comput. Sci.*, 360(1-3):209–227, 2006.
19. Ashwin Machanavajjhala, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. In *Proc. of ICDE*, pages 277–286. IEEE, 2008.
20. Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *Proc. of S&P*, pages 173–187. IEEE, 2009.
21. Jason Reed and Benjamin C. Pierce. Distance makes the types grow stronger: a calculus for differential privacy. In *Proc. of ICFP*, pages 157–168. ACM, 2010.
22. Peter Y. A. Ryan and Steve A. Schneider. Process algebra and non-interference. *Journal of Computer Security*, 9(1/2):75–103, 2001.
23. Geoffrey Smith. Probabilistic noninterference through weak probabilistic bisimulation. In *CSFW*, pages 3–13, 2003.
24. Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285—309, 1955.
25. Michael C. Tschantz, Dilsun Kaynar, and Anupam Datta. Formal verification of differential privacy for interactive systems (extended abstract). *ENTCS*, 276:61–79, sep 2011.
26. Lili Xu. Modular reasoning about differential privacy in a probabilistic process calculus. In *TGC*, pages 198–212, 2012.

A Appendix

Proofs are shown according to their orders in the main text.

A.1 Proof of Proposition 1

m^D is a pseudometric, that is:

1. (reflexivity) $m^D(s,s) = 0$
2. (symmetry) $m^D(s_1,s_2) = m^D(s_2,s_1)$
3. (triangle inequality) $m^D(s_1,s_3) \leq m^D(s_1,s_2) + m^D(s_2,s_3)$

Proof. 1. For reflexivity, it is enough to show that the identity relation over the set S of states of $\mathcal{A}$, that is the relation $Id_S = \{(s,s,0)|s \in S\}$, is an 0-accumulative bisimulation. This is easy.

2. For symmetry, assume that $(s_1,s_2,0)$ is in a ϵ-accumulative bisimulation $\mathcal{R}$, we will show that $\mathcal{R}' = \{(s_2',s_1',c)|(s_1',s_2',c) \in \mathcal{R}\}$ is a ϵ-accumulative bisimulation, thus we have $m^D(s_2,s_1) \leq \epsilon$.
   - It is easy to see that $(s_2,s_1,0) \in \mathcal{R}'$, because $(s_1,s_2,0) \in \mathcal{R}$.
   - For $(s_2',s_1',c) \in \mathcal{R}'$, if $s_2' \stackrel{a}{\Longrightarrow} \mu_2$, we must show that there exists a transition from s_1': $s_1' \stackrel{a}{\Longrightarrow} \mu_1$ and $\mu_2\mathcal{L}^D(\mathcal{R}',c)\mu_1$. Since $(s_1',s_2',c) \in \mathcal{R}$, there exists a transition from s_1' such that $s_1' \stackrel{a}{\Longrightarrow} \mu_1$ and $\mu_1\mathcal{L}^D(\mathcal{R},c)\mu_2$. According to the definition of D-approximate lifting, there exist a bijection $\beta : supp(\mu_1) \longrightarrow supp(\mu_2)$, such that for all s_1'' in $supp(\mu_1)$, $s_2'' = \beta(s_1'')$, $(s_1'',s_2'',c+\sigma) \in \mathcal{R}$ where $\sigma = \max_{s_1'' \in supp(\mu_1)} |\ln \frac{\mu_1(s_1'')}{\mu'(s_2'')}|$. Then $\mu_2\mathcal{L}^D(\mathcal{R}',c)\mu_1$ holds, because we have the inverse of the bijection β satisfying $s_1'' = \beta^{-1}(s_2'')$, and $(s_2'',s_1'',c+\sigma) \in \mathcal{R}'$.
   - For the other direction, it is analogous to the above case.

3. For transitivity, assume that $(s_1,s_2,0)$ is in the ϵ_1-accumulative bisimulation $\mathcal{R}_1 \subseteq S \times S \times [0,\epsilon_1]$, $(s_2,s_3,0)$ is in the ϵ_2-accumulative bisimulation $\mathcal{R}_2 \subseteq S \times S \times [0,\epsilon_2]$. we mush show that their relational composition $\mathcal{R}_1\mathcal{R}_2 \subseteq S \times S \times [0,\epsilon_1+\epsilon_2]$:

$$\{(s_1',s_3',c)|\exists s_2', c_1, c_2.(s_1',s_2',c_1) \in \mathcal{R}_1 \wedge (s_2',s_3',c_2) \in \mathcal{R}_2 \wedge c \leq c_1 + c_2\}$$

   is a $\epsilon_1 + \epsilon_2$-accumulative bisimulation.
   - It is easy to see that $(s_1,s_3,0) \in \mathcal{R}_1\mathcal{R}_2$, because $(s_1,s_2,0) \in \mathcal{R}_1$ and $(s_2,s_3,0) \in \mathcal{R}_2$.
   - for $(s_1',s_3',c) \in \mathcal{R}_1\mathcal{R}_2$, if $s_1' \stackrel{a}{\Longrightarrow} \mu_1$, we must show that there exists a transition from s_3': $s_3' \stackrel{a}{\Longrightarrow} \mu_3$ and $\mu_1\mathcal{L}^D(\mathcal{R}_1\mathcal{R}_2,c)\mu_3$. Since there exist s_2', c_1, c_2 such that $(s_1',s_2',c_1) \in \mathcal{R}_1$ and $(s_2',s_3',c_2) \in \mathcal{R}_2$ and $c \leq c_1 + c_2$, there exist also a transition $s_2' \stackrel{a}{\Longrightarrow} \mu_2$ and $\mu_1\mathcal{L}^D(\mathcal{R}_1,c_1)\mu_2$, and hence a transition $s_3' \stackrel{a}{\Longrightarrow} \mu_3$ and $\mu_2\mathcal{L}^D(\mathcal{R}_2,c_2)\mu_3$. By the definition of D-approximate lifting, there exists a bijection $\beta_1 : supp(\mu_1) \longrightarrow supp(\mu_2)$, s.t. for all s_1'' in $supp(\mu_1)$, $s_2'' = \beta_1(s_1'')$ and

$$(s_1'',s_2'',c_1+\sigma_1) \in \mathcal{R}_1 \text{ where } \sigma_1 = \max_{s_1'' \in supp(\mu_1)} |\ln \frac{\mu_1(s_1'')}{\mu_2(s_2'')}|.$$

There exists also a bijection $\beta_2 : supp(\mu_2) \longrightarrow supp(\mu_3)$, s.t. for all s_2'' in $supp(\mu_2)$, $s_3'' = \beta_2(s_2'')$ and

$$(s_2'', s_3'', c_2 + \sigma_2) \in \mathcal{R}_2 \text{ where } \sigma_2 = \max_{s_2'' \in supp(\mu_2)} |\ln \frac{\mu_2(s_2'')}{\mu_3(s_3'')}|.$$

It holds that $\mu_1 \mathcal{L}^D(\mathcal{R}_1\mathcal{R}_2, c)\mu_3$, because of the composition $\beta_1\beta_2$ satisfying $\beta_1\beta_2 : supp(\mu_1) \longrightarrow supp(\mu_3)$, s.t. for all s_1'' in $supp(\mu_1)$, $s_3'' = \beta_2(\beta_1(s_1''))$ and

$$(s_1'', s_3'', c + \sigma') \in \mathcal{R}_1\mathcal{R}_2 \text{ where } \sigma' = \max_{s_1'' \in supp(\mu_1)} |\ln \frac{\mu_1(s_1'')}{\mu_3(s_3'')}|$$

and $c + \sigma' \leq c_1 + \sigma_1 + c_2 + \sigma_2$.

- For the other direction, it is analogous to the above case.

□

A.2 Proof of Proposition 2

m^A is a pseudometric, that is:

1. (reflexivity) $m^A(s, s) = 0$
2. (symmetry) $m^A(s_1, s_2) = m^A(s_2, s_1)$
3. (triangle inequality) $m^A(s_1, s_3) \leq m^A(s_1, s_2) + m^A(s_2, s_3)$

Proof. 1. For reflexivity, it is enough to show that the identity relation over the set S of states of $\mathcal{A}$, that is the relation $Id_S = \{(s, s, 0) | s \in S\}$, is an 0-amortised bisimulation. This is easy.

2. For symmetry, assume that $(s_1, s_2, 0)$ is in a ϵ-amortised bisimulation $\mathcal{R}$, we will show that $\mathcal{R}' = \{(s_2', s_1', c) | (s_1', s_2', -c) \in \mathcal{R}\}$ is a ϵ-amortised bisimulation, thus we have $m^A(s_2, s_1) \leq \epsilon$.
   - It is easy to see that $(s_2, s_1, 0) \in \mathcal{R}'$, because $(s_1, s_2, 0) \in \mathcal{R}$.
   - for $(s_2', s_1', c) \in \mathcal{R}'$, if $s_2' \stackrel{a}{\Longrightarrow} \mu_2$, we must show that there exists a transition from s_1': $s_1' \stackrel{a}{\Longrightarrow} \mu_1$ and $\mu_2 \mathcal{L}^A(\mathcal{R}', c)\mu_1$. Since $(s_1', s_2', -c) \in \mathcal{R}$, there exists a transition from s_1' such that $s_1' \stackrel{a}{\Longrightarrow} \mu_1$ and $\mu_1 \mathcal{L}^A(\mathcal{R}, -c)\mu_2$. According to the definition of A-approximate lifting, there is a bijection $\beta : supp(\mu_1) \longrightarrow supp(\mu_2)$, s.t. for all s_1'' in $supp(\mu_1)$, $s_2'' = \beta(s_1'')$ and $(s_1'', s_2'', -c + \ln \mu_1(s_1'') - \ln \mu_2(s_2'')) \in \mathcal{R}$. Then $\mu_2 \mathcal{L}^A(\mathcal{R}', c)\mu_1$ holds, because we have the inverse of the bijection β satisfying $s_1'' = \beta^{-1}(s_2'')$, and $(s_2'', s_1'', c + \ln \mu_2(s_2'') - \ln \mu_1(s_1'')) \in \mathcal{R}'$.
   - For the other direction, it is analogous to the above case.
3. For transitivity, let $(s_1, s_2, 0)$ be in the ϵ_1-amortised bisimulation $\mathcal{R}_1 \subseteq S \times S \times [-\epsilon_1, \epsilon_1]$, $(s_2, s_3, 0)$ be in the ϵ_2-amortised bisimulation $\mathcal{R}_2 \subseteq S \times S \times [-\epsilon_2, \epsilon_2]$. we mush show that their relational composition $\mathcal{R}_1\mathcal{R}_2 \subseteq S \times S \times [-\epsilon_1 - \epsilon_2, \epsilon_1 + \epsilon_2]$:

$$\{(s_1', s_3', c) | \exists s_2', c_1, c_2.(s_1', s_2', c_1) \in \mathcal{R}_1 \wedge (s_2', s_3', c_2) \in \mathcal{R}_2 \wedge c_1 + c_2 = c\}$$

is a $\epsilon_1 + \epsilon_2$-amortised bisimulation.

- It is easy to see that $(s_1, s_3, 0) \in \mathcal{R}_1\mathcal{R}_2$, because $(s_1, s_2, 0) \in \mathcal{R}_1$ and $(s_2, s_3, 0) \in \mathcal{R}_2$.
- for $(s_1', s_3', c) \in \mathcal{R}_1\mathcal{R}_2$, if $s_1' \stackrel{a}{\Longrightarrow} \mu_1$, we must show that there exists a transition from s_3': $s_3' \stackrel{a}{\Longrightarrow} \mu_3$ and $\mu_1 \mathcal{L}^A(\mathcal{R}_1\mathcal{R}_2, c)\mu_3$. Since there exist s_2', c_1, c_2 such that $(s_1', s_2', c_1) \in \mathcal{R}_1$ and $(s_2', s_3', c_2) \in \mathcal{R}_2$ and $c_1 + c_2 = c$, there exist also a transition $s_2' \stackrel{a}{\Longrightarrow} \mu_2$ and $\mu_1 \mathcal{L}^A(\mathcal{R}_1, c_1)\mu_2$, and hence a transition $s_3' \stackrel{a}{\Longrightarrow} \mu_3$ and $\mu_2 \mathcal{L}^A(\mathcal{R}_2, c_2)\mu_3$. By the definition of A-approximate lifting, there is a bijection $\beta_1 : supp(\mu_1) \longrightarrow supp(\mu_2)$, s.t. for all s_1'' in $supp(\mu_1)$,

  $$s_2'' = \beta_1(s_1'') \text{ and } (s_1'', s_2'', c_1 + \ln \mu_1(s_1'') - \ln \mu_2(s_2'')) \in \mathcal{R}_1.$$

  There is also a bijection $\beta_2 : supp(\mu_2) \longrightarrow supp(\mu_3)$, s.t. for all s_2'' in $supp(\mu_2)$,

  $$s_3'' = \beta_2(s_2'') \text{ and } (s_2'', s_3'', c_2 + \ln \mu_2(s_2'') - \ln \mu_3(s_3'')) \in \mathcal{R}_2.$$

  It holds that $\mu_1 \mathcal{L}^A(\mathcal{R}_1\mathcal{R}_2, c)\mu_3$, because we have the composition $\beta_1\beta_2$ satisfying $\beta_1\beta_2 : supp(\mu_1) \longrightarrow supp(\mu_3)$, s.t. for all s_1'' in $supp(\mu_1)$,

  $$s_3'' = \beta_2(\beta_1(s_1'')) \text{ and } (s_1'', s_3'', c + \ln \mu_1(s_1'') - \ln \mu_3(s_3'')) \in \mathcal{R}_1\mathcal{R}_2.$$

- For the other direction, it is analogous to the above case.

□

A.3 Proof of Lemma 2

Given a PA $\mathcal{A}$, let $\mathcal{R}$ be an ϵ-amortised bisimulation, $c \in [-\epsilon, \epsilon]$, let ζ be an admissible scheduler, $\boldsymbol{t}$ be a finite trace, α_1, α_2 two finite executions of $\mathcal{A}$. If $(lstate(\alpha_1), lstate(\alpha_2), c) \in \mathcal{R}$, then

$$\frac{1}{e^{\epsilon+c}} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^{\epsilon-c}$$

Proof. We prove by induction on the length of trace $\boldsymbol{t}$: $|\boldsymbol{t}|$.

1. $|\boldsymbol{t}| = 0$: According to equation (1), for any scheduler ζ, $\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] = \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] = 1$.
2. IH: For any two executions α_1 and α_2 of $\mathcal{A}$, let $s_1 = lstate(\alpha_1)$ and $s_2 = lstate(\alpha_2)$. $(s_1, s_2, c) \in \mathcal{R}$ implies that for any admissible scheduler ζ, trace $\boldsymbol{t}'$ where $|\boldsymbol{t}'| \leq L$:
   $$\frac{1}{e^{\epsilon+c}} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}']}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}']} \leq e^{\epsilon-c}$$
3. We have to show that for any admissible scheduler ζ, trace $\boldsymbol{t}$ with $|\boldsymbol{t}| = L + 1$, $(s_1, s_2, c) \in \mathcal{R}$ implies

   $$\frac{1}{e^{\epsilon+c}} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^{\epsilon-c}$$

   Assume that $\boldsymbol{t} = a^\frown \boldsymbol{t}'$. We prove first the right-hand part $\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] \leq e^{\epsilon-c} * \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]$. According to equation (1), two cases must be considered:

- Case $act(\zeta(\alpha_1)) \neq a$. Then $\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] = 0$. Since ζ is admissible, it schedules for α_2 a transition consistent with α_1, namely, not a transition labeled by a either. Thus $\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] = 0$, the inequality is satisfied.
- Case $\zeta(\alpha_1) = s_1 \stackrel{a}{\Longrightarrow} \mu_1$. So,

$$\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] = \textstyle\sum_{s_i \in supp(\mu_1)} \mu_1(s_i) \Pr_\zeta[\alpha_1 a s_i \rhd \boldsymbol{t}']$$

Since $(s_1, s_2, c) \in \mathcal{R}$, there must be also a transition from s_2 such that $s_2 \stackrel{a}{\Longrightarrow} \mu_2$ and $\mu_1 \mathcal{L}^A(\mathcal{R}, c) \mu_2$. Since ζ is admissible, $\zeta(\alpha_2) = s_2 \stackrel{a}{\Longrightarrow} \mu_2$. We use t_i to range over elements in $supp(\mu_2)$. Thus,

$$\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] = \textstyle\sum_{t_i \in supp(\mu_2)} \mu_2(t_i) \Pr_\zeta[\alpha_2 a t_i \rhd \boldsymbol{t}']$$

Since $\mu_1 \mathcal{L}^A(\mathcal{R}, c) \mu_2$, there is a bijection $\beta : supp(\mu_1) \longrightarrow supp(\mu_2)$, s.t. for any $s_i \in supp(\mu_1)$, there is a state $t_i \in supp(\mu_2)$, $t_i = \beta(s_i)$ and $(s_i, t_i, c + \ln \mu_1(s_i) - \ln \mu_2(t_i)) \in \mathcal{R}$. Apply the inductive hypothesis to $\alpha_1 a s_i, \alpha_2 a t_i$ and $\boldsymbol{t}'$, we get that:

$$\Pr_\zeta[\alpha_1 a s_i \rhd \boldsymbol{t}'] \leq e^{\epsilon - (c + \ln \mu_1(s_i) - \ln \mu_2(t_i))} * \Pr_\zeta[\alpha_2 a t_i \rhd \boldsymbol{t}'] \tag{2}$$

Thus,

$$\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] \tag{3}$$

$$= \sum_{s_i \in supp(\mu_1)} \mu_1(s_i) \Pr_\zeta[\alpha_1 a s_i \rhd \boldsymbol{t}'] \tag{4}$$

$$\leq \sum_{s_i \in supp(\mu_1)} \mu_1(s_i) e^{\epsilon - (c + \ln \mu_1(s_i) - \ln \mu_2(\beta(s_i)))} \Pr_\zeta[\alpha_2 a \beta(s_i) \rhd \boldsymbol{t}'] \tag{5}$$

$$= \sum_{s_i \in supp(\mu_1)} \mu_1(s_i) * \frac{\mu_2(\beta(s_i))}{\mu_1(s_i)} * e^{\epsilon - c} * \Pr_\zeta[\alpha_2 a \beta(s_i) \rhd \boldsymbol{t}'] \tag{6}$$

$$= \sum_{t_i \in supp(\mu_2)} \mu_2(t_i) * e^{\epsilon - c} * \Pr_\zeta[\alpha_2 a t_i \rhd \boldsymbol{t}'] \tag{7}$$

$$= e^{\epsilon - c} \sum_{t_i \in supp(\mu_2)} \mu_2(t_i) \Pr_\zeta[\alpha_2 a t_i \rhd \boldsymbol{t}'] \tag{8}$$

$$= e^{\epsilon - c} * \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] \tag{9}$$

which completes the proof of right-hand part. Lines (4) and (9) follow from the equation (1). Line (5) follow from the inductive hypothesis, i.e. Line (2).
For the left-hand part $\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] \leq e^{\epsilon + c} * \Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]$, exchange the roles of s_1 and s_2. use β^{-1} instead of β, and all the rest is analogous. □

A.4 Proof of Theorem 2

A concurrent system $\mathcal{A}$ is ϵ-differentially private if $m^A(\mathcal{A}(u), \mathcal{A}(u')) \leq \epsilon$ for all $u \sim u'$.

Proof. Since $m^A(\mathcal{A}(u), \mathcal{A}(u')) \leq \epsilon$ for all $u \sim u'$, by the definition of m^A, there exists a ϵ-amortised bisimulation $\mathcal{R}$ such that $(\mathcal{A}(u), \mathcal{A}(u'), 0) \in \mathcal{R}$. By Lemma 2, for any admissible scheduler ζ, any finite trace $\boldsymbol{t}$:

$$\frac{1}{e^\epsilon} \leq \frac{\Pr_\zeta[\mathcal{A}(u) \rhd \boldsymbol{t}]}{\Pr_\zeta[\mathcal{A}(u') \rhd \boldsymbol{t}]} \leq e^\epsilon$$

Thus, $\mathcal{A}$ is ϵ-differentially private. □

A.5 Middle steps of the transformation of the linear-fractional program

Here shows the detailed transformation of the following linear-fractional program to its linear counterpart and the dual program.

$$\begin{aligned} \text{maximize} \quad & \frac{\sum_i \mu(s_i)x_i}{\sum_i \mu'(s_i)x_i} \\ \text{subject to:} \quad & \forall i.\ 0 \leq x_i \leq 1 \\ & \forall i, j.\ x_i \leq e^{m(s_i,s_j)}x_j. \end{aligned}$$

Following the techniques in [5], we extend the dimensions of the feasible region by adding new decision variables y_i for $i \in [1, |S|]$. The extension does not affect the optimal value. This is justified by the new constraints ensuring that in fact $x_i = y_i$ for $i \in [1, |S|]$ (because $m(s_i, s_i) = 0$).

$$\begin{aligned} \text{maximize} \quad & \frac{\sum_i \mu(s_i)x_i}{\sum_j \mu'(s_j)y_j} \\ \text{subject to:} \quad & \forall i.\ 0 \leq x_i, y_i \leq 1 \\ & \forall i, j.\ x_i - e^{m(s_i,s_j)}y_j \leq 0 \\ & \forall i,\ y_i - x_i \leq 0. \end{aligned}$$

Let $\alpha_i = \dfrac{x_i}{\sum_j \mu'(s_j)y_j}$, $\beta_i = \dfrac{y_i}{\sum_j \mu'(s_j)y_j}$ and $t = \dfrac{1}{\sum_j \mu'(s_j)y_j}$. The above linear-fractional problem can be transformed to the equivalent linear program.

$$\begin{array}{ll} \text{maximize} & \sum_i \mu(s_i)\alpha_i \\ \\ \text{subject to:} & \forall i.\ 0 \leq \alpha_i, \beta_i \leq t \\ & \forall i, j.\ \alpha_i - e^{m(s_i,s_j)}\beta_j \leq 0 \\ & \forall i,\ \beta_i - \alpha_i \leq 0 \\ & \sum_i \mu'(s_i)\beta_i = 1. \end{array}$$

Dualizing the above (primal) problem yields:

$$\begin{array}{ll} \text{minmize} & z \\ \text{subject to:} & \forall i. \sum_j l_{ij} + a_i - r_i \geq \mu(s_i) \\ & \forall j. \sum_i l_{ij} e^{m(s_i,s_j)} - b_j - r_j \leq z \cdot \mu'(s_j) \\ & \sum_i a_i + \sum_i b_i \leq 0 \\ & \forall i, j.\ l_{ij}, a_i, b_i, r_i \geq 0 \end{array}$$

where the possible values for a_i and b_i can only be 0. Thus, the dual problem becomes:

$$\begin{array}{ll} \text{minmize} & z \\ \text{subject to:} & \forall i. \sum_j l_{ij} - r_i \geq \mu(s_i) \\ & \forall j. \sum_i l_{ij} e^{m(s_i,s_j)} - r_j \leq z \cdot \mu'(s_j) \\ & \forall i, j.\ l_{ij}, r_i \geq 0. \end{array}$$

which is equivalent to the following program where the first constraints are equations:

$$\begin{array}{ll} \text{minmize} & z \\ \text{subject to:} & \forall i. \sum_j l_{ij} - r_i = \mu(s_i) \\ & \forall j. \sum_i l_{ij} e^{m(s_i,s_j)} - r_j \leq z \cdot \mu'(s_j) \\ & \forall i, j.\ l_{ij}, r_i \geq 0. \end{array}$$

A.6 Proof of Lemma 3

Def. 6 defines $m(\mu, \mu')$ as the solution to the following program:

$$\text{maximize} \qquad |\ln \frac{\sum_i \mu(s_i)x_i}{\sum_i \mu'(s_i)x_i}| \tag{10}$$

We prove Lemma 3 as follows:

1. Let $m \in \mathcal{M}$. $m(\mu, \mu) = 0$.
2. (Symmetry.) Let $m \in \mathcal{M}$. $m(\mu, \mu') = m(\mu', \mu)$.
3. (Triangle inequality.) Let $m \in \mathcal{M}$. Then, $(\forall \mu_1, \mu_2, \mu_3)\, m(\mu_1, \mu_3) \leq m(\mu_1, \mu_2) + m(\mu_2, \mu_3)$.
4. Let $m, m' \in \mathcal{M}$ such that $m \preceq m'$. Then, for all distributions on states μ, μ', $m(\mu, \mu') \geq m'(\mu, \mu')$.

Proof. 1. Straightforwardly, the numerator and the denominator in the program of Definition 6 turn out to be equivalent. Thus $m(\mu, \mu) = \ln 1 = 0$.

2. It is easy to see the symmetry property, because $|\ln \sum_i \mu(s_i)x_i - \ln \sum_i \mu'(s_i)x_i| = |\ln \sum_i \mu'(s_i)x_i - \ln \sum_i \mu(s_i)x_i|$.
3. To prove the triangle inequality, given distributions μ_1, μ_2, μ_3,

$$\begin{aligned}
|\ln \frac{\sum_i \mu_1(s_i)x_i}{\sum_i \mu_3(s_i)x_i}| &= |\ln(\frac{\sum_i \mu_1(s_i)x_i}{\sum_i \mu_2(s_i)x_i} \cdot \frac{\sum_i \mu_2(s_i)x_i}{\sum_i \mu_3(s_i)x_i})| \\
&= |\ln \frac{\sum_i \mu_1(s_i)x_i}{\sum_i \mu_2(s_i)x_i} + \ln \frac{\sum_i \mu_2(s_i)x_i}{\sum_i \mu_3(s_i)x_i}| \\
&\leq |\ln \frac{\sum_i \mu_1(s_i)x_i}{\sum_i \mu_2(s_i)x_i}| + |\ln \frac{\sum_i \mu_2(s_i)x_i}{\sum_i \mu_3(s_i)x_i}|
\end{aligned}$$

   Taking the maximum over the x_i for the left side, we get $m(\mu_1, \mu_3) \leq m(\mu_1, \mu_2) + m(\mu_2, \mu_3)$.
4. For the last item, consider the primal program (10), note that every solution to the program defining $m'(\mu_1, \mu_2)$ is also a solution to the program defining $m(\mu_1, \mu_2)$. So, the maximum value $m'(\mu_1, \mu_2) \leq m(\mu_1, \mu_2)$.

□

A.7 Proof of Lemma 4

Given a PA $\mathcal{A}$, let ζ be an admissible scheduler, let $\boldsymbol{t}$ be a finite trace, and let α_1, α_2 be two states in $\mathcal{A}_\zeta$. If $m^K(lstate(\alpha_1), lstate(\alpha_2)) \leq \epsilon$ then

$$\frac{1}{e^\epsilon} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^\epsilon$$

Proof. We prove by induction on the length of trace $\boldsymbol{t}$: $|\boldsymbol{t}|$.

1. $|\boldsymbol{t}| = 0$: According to equation (1), $\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] = \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] = 1$.
2. IH: For any two states α_1 and α_2 in $\mathcal{A}_\zeta$, let $s_1 = lstate(\alpha_1)$ and $s_2 = lstate(\alpha_2)$, $m^K(s_1, s_2) \leq \epsilon$ implies that for any trace $\boldsymbol{t}'$ where $|\boldsymbol{t}'| \leq L$,

$$\frac{1}{e^\epsilon} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}']}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}']} \leq e^\epsilon$$

Because $m^K(s,s) = 0$, we can easily deduce from the IH that

$$\Pr_\zeta[\alpha \rhd \boldsymbol{t}'] = \Pr_\zeta[s \rhd \boldsymbol{t}'] \tag{11}$$

in which $s = lstate(\alpha)$.

3. We have to show that for any trace $\boldsymbol{t}$ with $|\boldsymbol{t}| = L+1$, $m^K(s_1,s_2) \leq \epsilon$ implies:

$$\frac{1}{e^\epsilon} \leq \frac{\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^\epsilon$$

Assume that $\boldsymbol{t} = a^\frown \boldsymbol{t}'$. According to equation (1), two cases must be considered:

- Case $act(\zeta(\alpha_1)) \neq a$. Then $\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] = 0$. Since ζ is an admissible scheduler, it does not schedule for α_2 the transition with label a either. Thus $\Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] = 0$, the inequality is satisfied.
- Case $\zeta(\alpha_1) = s_1 \stackrel{a}{\Longrightarrow} \mu_1$. So, by (11) we have

$$\begin{aligned} \Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] &= \textstyle\sum_{s_i \in supp(\mu_1)} \mu_1(s_i) \Pr_\zeta[\alpha_1 a s_i \rhd \boldsymbol{t}'] \\ &= \textstyle\sum_i \mu_1(s_i) \Pr_\zeta[s_i \rhd \boldsymbol{t}'] \end{aligned} \tag{12}$$

Since ζ is admissible, $m^K(s_1,s_2) \leq \epsilon$ which is the maximum fixed-point of F, there must be also a transition from s_2 such that $s_2 \stackrel{a}{\Longrightarrow} \mu_2$, $m^K(\mu_1,\mu_2) \leq \epsilon$, and $\zeta(\alpha_2) = s_2 \stackrel{a}{\Longrightarrow} \mu_2$. So, by (11) we have analogously

$$\begin{aligned} \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] &= \textstyle\sum_{s_i \in supp(\mu_2)} \mu_2(s_i) \Pr_\zeta[\alpha_2 a s_i \rhd \boldsymbol{t}'] \\ &= \textstyle\sum_i \mu_2(s_i) \Pr_\zeta[s_i \rhd \boldsymbol{t}'] \end{aligned} \tag{13}$$

Since $m^K(\mu_1,\mu_2) \leq \epsilon$, by Def. 6, it follows that:

$$\begin{aligned} \max |\ln \frac{\sum_i \mu_1(s_i) x_i}{\sum_i \mu_2(s_i) x_i}| &\leq \epsilon \\ \text{subject to:} \quad \forall i. 0 \leq x_i \leq 1 & \\ \forall i,j. \frac{x_i}{x_j} &\leq e^{m^K(s_i,s_j)} \end{aligned} \tag{14}$$

Let $x_i = \Pr_\zeta[s_i \rhd \boldsymbol{t}']$. Clearly, it holds that $0 \leq \Pr_\zeta[s_i \rhd \boldsymbol{t}'] \leq 1$. In addition, applying the inductive hypothesis to s_i, s_j and $\boldsymbol{t}'$, we get:

$$\frac{1}{e^{m^K(s_i,s_j)}} \leq \frac{\Pr_\zeta[s_i \rhd \boldsymbol{t}']}{\Pr_\zeta[s_j \rhd \boldsymbol{t}']} \leq e^{m^K(s_i,s_j)}$$

Thus by equation (14), we have:

$$|\ln \frac{\sum_i \mu_1(s_i) \Pr_\zeta[s_i \rhd \boldsymbol{t}']}{\sum_i \mu_2(s_i) \Pr_\zeta[s_i \rhd \boldsymbol{t}']}| \leq \epsilon$$

$$-\epsilon \leq \ln \frac{\sum_i \mu_1(s_i)\mathrm{Pr}_\zeta[s_i \rhd \boldsymbol{t}']}{\sum_i \mu_2(s_i)\mathrm{Pr}_\zeta[s_i \rhd \boldsymbol{t}']} \leq \epsilon$$

$$\frac{1}{e^\epsilon} \leq \frac{\sum_i \mu_1(s_i)\mathrm{Pr}_\zeta[s_i \rhd \boldsymbol{t}']}{\sum_i \mu_2(s_i)\mathrm{Pr}_\zeta[s_i \rhd \boldsymbol{t}']} \leq e^\epsilon$$

By equations (12) and (13), we obtain:

$$\frac{1}{e^\epsilon} \leq \frac{\mathrm{Pr}_\zeta[\alpha_1 \rhd \boldsymbol{t}]}{\mathrm{Pr}_\zeta[\alpha_2 \rhd \boldsymbol{t}]} \leq e^\epsilon$$

as required. □

A.8 Proof of Lemma 5

Let m be the pseudometric based on the standard Kantorovich metric proposed in [11].

Given a PA $\mathcal{A}$, let ζ be an admissible scheduler, let $\boldsymbol{t}$ be a finite trace, and let α_1, α_2 be two states in $\mathcal{A}_\zeta$. If $m(lstate(\alpha_1), lstate(\alpha_2)) \leq \epsilon$ then

$$|\,\mathrm{Pr}_\zeta[\alpha_1 \rhd \boldsymbol{t}] - \mathrm{Pr}_\zeta[\alpha_2 \rhd \boldsymbol{t}]\,| \leq \epsilon.$$

Proof. We prove by induction on the length of trace $\boldsymbol{t}$: $|\boldsymbol{t}|$.

1. $|\boldsymbol{t}| = 0$: According to equation (1), $\mathrm{Pr}_\zeta[\alpha_1 \rhd \boldsymbol{t}] = \mathrm{Pr}_\zeta[\alpha_2 \rhd \boldsymbol{t}] = 1$.
2. IH: For any two states α_1 and α_2 in $\mathcal{A}_\zeta$, let $s_1 = lstate(\alpha_1)$ and $s_2 = lstate(\alpha_2)$, $m(s_1, s_2) \leq \epsilon$ implies that for any trace $\boldsymbol{t}'$ where $|\boldsymbol{t}'| \leq L$,

   $$|\,\mathrm{Pr}_\zeta[\alpha_1 \rhd \boldsymbol{t}'] - \mathrm{Pr}_\zeta[\alpha_2 \rhd \boldsymbol{t}']\,| \leq \epsilon$$

   Because $m(s, s) = 0$, we can easily deduce from the IH that

   $$\mathrm{Pr}_\zeta[\alpha \rhd \boldsymbol{t}'] = \mathrm{Pr}_\zeta[s \rhd \boldsymbol{t}'] \tag{15}$$

   in which $s = lstate(\alpha)$.
3. We have to show that for any trace $\boldsymbol{t}$ with $|\boldsymbol{t}| = L + 1$, $m(s_1, s_2) \leq \epsilon$ implies:

   $$|\,\mathrm{Pr}_\zeta[\alpha_1 \rhd \boldsymbol{t}] - \mathrm{Pr}_\zeta[\alpha_2 \rhd \boldsymbol{t}]\,| \leq \epsilon.$$

   Assume that $\boldsymbol{t} = a^\frown \boldsymbol{t}'$. According to equation (1), two cases must be considered:
   - Case $act(\zeta(\alpha_1)) \neq a$. Then $\mathrm{Pr}_\zeta[\alpha_1 \rhd \boldsymbol{t}] = 0$. Since ζ is an admissible scheduler, it does not schedule for α_2 the transition with label a either. Thus $\mathrm{Pr}_\zeta[\alpha_2 \rhd \boldsymbol{t}] = 0$, the inequality is satisfied.
   - Case $\zeta(\alpha_1) = s_1 \stackrel{a}{\Longrightarrow} \mu_1$. So, by (15) we have

   $$\begin{aligned} \mathrm{Pr}_\zeta[\alpha_1 \rhd \boldsymbol{t}] &= \textstyle\sum_{s_i \in supp(\mu_1)} \mu_1(s_i)\mathrm{Pr}_\zeta[\alpha_1 a s_i \rhd \boldsymbol{t}'] \\ &= \textstyle\sum_i \mu_1(s_i)\mathrm{Pr}_\zeta[s_i \rhd \boldsymbol{t}'] \end{aligned} \tag{16}$$

Since $m(s_1, s_2) \leq \epsilon$ and ζ is admissible, by the definition of the pseudometric in [11], there must be also a transition from s_2 such that $s_2 \stackrel{a}{\Longrightarrow} \mu_2$, $m(\mu_1, \mu_2) \leq \epsilon$, and $\zeta(\alpha_2) = s_2 \stackrel{a}{\Longrightarrow} \mu_2$. So, by (15) we have analogously

$$\begin{aligned} \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}] &= \textstyle\sum_{s_i \in supp(\mu_2)} \mu_2(s_i) \Pr_\zeta[\alpha_2 a s_i \rhd \boldsymbol{t}'] \\ &= \textstyle\sum_i \mu_2(s_i) \Pr_\zeta[s_i \rhd \boldsymbol{t}'] \end{aligned} \tag{17}$$

Since $m(\mu_1, \mu_2) \leq \epsilon$, it follows that:

$$\max |\sum_i \mu_1(s_i) x_i - \sum_i \mu_2(s_i) x_i| \leq \epsilon \tag{18}$$

$$\begin{aligned} \text{subject to:} \quad & \forall i. 0 \leq x_i \leq 1 \\ & \forall i, j. x_i - x_j \leq m(s_i, s_j) \end{aligned}$$

Let $x_i = \Pr_\zeta[s_i \rhd \boldsymbol{t}']$. Clearly, it holds that $0 \leq \Pr_\zeta[s_i \rhd \boldsymbol{t}'] \leq 1$. In addition, applying the inductive hypothesis to s_i, s_j and $\boldsymbol{t}'$, we get:

$$|\Pr_\zeta[s_i \rhd \boldsymbol{t}'] - \Pr_\zeta[s_j \rhd \boldsymbol{t}']| \leq m(s_i, s_j)$$

Thus by equation (18), we have:

$$|\sum_i \mu_1(s_i) \Pr_\zeta[s_i \rhd \boldsymbol{t}'] - \sum_i \mu_2(s_i) \Pr_\zeta[s_i \rhd \boldsymbol{t}']| \leq \epsilon$$

By equations (16) and (17), we obtain:

$$|\Pr_\zeta[\alpha_1 \rhd \boldsymbol{t}] - \Pr_\zeta[\alpha_2 \rhd \boldsymbol{t}]| \leq \epsilon$$

as required. □

A.9 Proof of Lemma 6

$m^D \preceq m^A$.

Proof. Assume that $\mathcal{R}^D \subseteq S \times S \times [0, \epsilon]$ is the ϵ-accumulative bisimulation such that $(s, t, 0) \in \mathcal{R}^D$. We define a relation $\mathcal{R}^A \subseteq S \times S \times [-\epsilon, \epsilon]$ from $\mathcal{R}^D$ as follows:

$$(s', t', c^A) \in \mathcal{R}^A \text{ iff } \exists c^D.(s', t', c^D) \in \mathcal{R}^D \wedge |c^A| \leq c^D \tag{19}$$

Now we prove that $\mathcal{R}^A$ is an ϵ-amortised bisimulation.

1. It is easy to see that $(s, t, 0) \in \mathcal{R}^A$, because $(s, t, 0) \in \mathcal{R}^D$.
2. Given $(s', t', c^A) \in \mathcal{R}^A$, if $s' \stackrel{a}{\Longrightarrow} \mu_1$, we must show that there exists a transition from t': $t' \stackrel{a}{\Longrightarrow} \mu_2$ and $\mu_1 \mathcal{L}^A(\mathcal{R}^A, c^A) \mu_2$. By (19) we know that there exists c^D such that $|c^A| \leq c^D$ and $(s', t', c^D) \in \mathcal{R}^D$. Thus there exists a transition from t' such that $t' \stackrel{a}{\Longrightarrow} \mu_2$ and $\mu_1 \mathcal{L}^D(\mathcal{R}^D, c^D) \mu_2$. According to the definition of D-approximate lifting, there exists a bijection $\beta : supp(\mu_1) \longrightarrow supp(\mu_2)$, s.t. for all s'' in $supp(\mu_1)$, $t'' = \beta(s'')$, $(s'', t'', c^D + \sigma) \in \mathcal{R}^D$ where $\sigma = \max_{s'' \in supp(\mu_1)} |\ln \frac{\mu_1(s'')}{\mu_2(t'')}|$. We have $|c^A + \ln \mu_1(s'') - \ln \mu_2(t'')| \leq c^D + \sigma$ and hence $(s'', t'', c^A + \ln \mu_1(s'') - \ln \mu_2(t'')) \in \mathcal{R}^A$ by (19). According to the definition of A-approximate lifting, it holds that $\mu_1 \mathcal{L}^A(\mathcal{R}^A, c^A) \mu_2$ as required.

3. For the other direction, it is analogous to the above case.

□

A.10 Proof of Lemma 7

$m^D \preceq m^K$.

Proof. We need to show that $m^D(s,t) \leq \epsilon$ implies $m^K(s,t) \leq \epsilon$. Since $m^D(s,t) \leq \epsilon$, there exists an ϵ-accumulative bisimulation $\mathcal{R}$ such that $(s,t,0) \in \mathcal{R}$. We show that if $(s,t,c) \in \mathcal{R}$ then $m^D(s,t) \leq \epsilon - c$. It is justified by proving that $\mathcal{R}' \subseteq S \times S \times [0, \epsilon - c]$ defined as $\{(s,t,c') \mid (s,t,c'+c) \in \mathcal{R}\}$ is an $(\epsilon - c)$-accumulative bisimulation.

For any $s \stackrel{a}{\Longrightarrow} \mu$, there exist $t \stackrel{a}{\Longrightarrow} \mu'$ and a bijection $\beta : supp(\mu) \to supp(\mu')$ such that for all s_i in $supp(\mu)$, $t_i = \beta(s_i)$ and $(s_i, t_i, \sigma) \in \mathcal{R}$ where $\sigma = \max_i |\ln \mu(s_i) - \ln \mu'(t_i)|$. We prove that the distance given by the multiplicative Kantorovich metric $m^D(\mu, \mu')$ is less than ϵ.

$$m^D(\mu,\mu') = \max |\ln \frac{\sum_i \mu(s_i)x_i}{\sum_i \mu'(t_i)y_i}|$$

where $|\ln \frac{x_i}{y_i}| \leq m^D(s_i,t_i)$. In particular, for $t_i = \beta(s_i)$, we have $|\ln \frac{x_i}{y_i}| \leq m^D(s_i,t_i) \leq \epsilon - \sigma$. Thus,

$$\begin{aligned}
m^D(\mu,\mu') &\leq \max |\ln \frac{\sum_i \mu(s_i)y_i e^{\epsilon-\sigma}}{\sum_i \mu'(t_i)y_i}| \\
&= \max |\ln \frac{e^{\epsilon}\sum_i \mu(s_i)y_i e^{-\sigma}}{\sum_i \mu'(t_i)y_i}| \\
&\leq \max |\ln \frac{e^{\epsilon}\sum_i \mu(s_i)y_i e^{\ln\mu'(t_i) - \ln\mu(s_i)}}{\sum_i \mu'(t_i)y_i}| \\
&= \max |\ln \frac{e^{\epsilon}\sum_i \mu'(t_i)y_i}{\sum_i \mu'(t_i)y_i}| \\
&= \epsilon
\end{aligned}$$

It satisfies that $F(m^D)(s,t) \leq m^D(s,t)$, showing that m^D is also a pre-fixed point of F. Henceforth the greatest fixed-point of F: $m^K(s,t) \leq \epsilon$. □

A.11 Computations for Example 4

Consider the processes s, t shown in Fig. 4. We compute their distance using the third pseudometric, showing that $m^K(s,t) = \ln 24$.

We denote $s \stackrel{a}{\Longrightarrow} \mu_1$, $t \stackrel{a}{\Longrightarrow} \mu_2$, $s_1 \stackrel{b}{\Longrightarrow} \eta_1$ and $t_1 \stackrel{b}{\Longrightarrow} \eta_2$. Compute

$$m(\eta_1,\eta_2) = \max |\ln \frac{0.2x_1 + 0.1x_2 + 0.7x_3}{0.2y_1 + 0.6y_2 + 0.2y_3}|$$

under the constraints: $\forall i \in \{1,2,3\}, 0 \leq x_i, y_i \leq 1$, $x_1 = y_1, x_2 = y_2, x_3 = y_3$, x_i and y_j are independent from each other if $i \neq j$. We obtain $m(\eta_1, \eta_2) = \ln 6$. Thus $m(s_1, t_1) = \ln 6$. Compute

$$m(\mu_1, \mu_2) = \max |\ln \frac{0.4x_1 + 0.6x_2}{0.1y_1 + 0.9y_2}|$$

under the constraints: $\forall i \in \{1,2\}, 0 \leq x_i, y_i \leq 1$, $x_1 \leq 6y_1, y_1 \leq 6x_1, x_2 = y_2$, x_i and y_j are independent from each other if $i \neq j$. We obtain $m(s,t) = m(\mu_1, \mu_2) = \ln 24$. Since there is no iteration, $m^K(s,t) = \ln 24$.

Next, we compute their distance by using the second pseudometric, showing that $m^A(s,t) = \ln 14$ which is finer than the former distance. Let S and T denote the state space of systems s and t, respectively. Let $\mathcal{R} \subseteq S \times T \times [-\ln 14, \ln 14]$. It is straightforward to check according to Def. 5 that the following relation is an amortised bisimulation between s and t.

$$\begin{aligned}\mathcal{R} = \{\ & (s,t,0),\\ & (s_1,t_1,\ln 4),\ \ (s_4,t_4,\ln \tfrac{2}{3}),\\ & (s_2,t_2,\ln 4),\ \ (s_3,t_3,\ln \tfrac{2}{3}),\\ & (s_5,s_5,\ln 14)\ \}\end{aligned}$$

It is easy to see that $m^A(s,t) = \ln 14$ since there does not exist an amortised bisimulation with a ϵ smaller that $\ln 14$.

A.12 Proof of Proposition 3: the first item

We adopt the notion of weak bisimilarity proposed in [11]. The "probability" from a state s to a subset of states via a trace with weak label a is defined by taking the supremum over all possible computations.

Definition 9. *Let $\mathcal{A}$ be a PA, $s \in S$, $E \subseteq S$. Then, the probability of going from s to E via a, denoted by $\mu(s,a,E)$, is defined as:*

$$\mu(s,a,E) = \sup\{\sum_{t \in E} \mu'(t) \mid s \stackrel{a}{\Longrightarrow} \mu'\}.$$

In [11], it has been proved that there exists a computation with root s that assigns the maximum probability to E, i.e. $\mu(s,a,E) = \sum_{t \in E} \mu'(t)$ for some $s \stackrel{a}{\Longrightarrow} \mu'$.

We consider equivalence relations on the set of states. Given an equivalence relation $\mathcal{R} \subseteq S \times S$, we say a set E is $\mathcal{R}$-closed if $E = \{s \mid \exists t \in E$ such that $t\mathcal{R}s\}$.

Definition 10. *An equivalence relation $\mathcal{R} \subseteq S \times S$ is a* weak bisimulation *if for all $s,t \in S$ such that $s\mathcal{R}t$ and all $\mathcal{R}$-closed $E \subseteq S$, we have:*

$$(\forall a \in A)[\mu(s,a,E) = \mu(t,a,E)].$$

There is a maximum weak bisimulation, namely weak bisimilarity*, denoted by $\approx$.*

$$s \approx t \Leftrightarrow m^K(s,t) = 0.$$

Proof. ($\Rightarrow$) Consider the pseudometric m, defined as $m(s,t) = 0$ if s and t are weak bisimilar, and ∞ otherwise. By Lemma 2.9 in [11], which states that given $s \approx t$, if $s \stackrel{a}{\Longrightarrow} \mu$ then there exists $t \stackrel{a}{\Longrightarrow} \mu'$ such that for all states s_i: $\mu([s_i]) = \mu'([s_i])$, where $[s_i] \in S/\approx$. Consider the primal program (10) determining the value of $m(\mu, \mu')$, note that $s_i \approx s_j$ implies $x_i = x_j$. Thus the summations in the objective function can be grouped together by the equivalence classes of S under $\approx$. It is now straightforward to see that $\sum_{[s_i] \in S/\approx} \mu([s_i])x_i = \sum_{[s_i] \in S/\approx} \mu'([s_i])x_i$ and hence $m(\mu, \mu') = 0$. For the bisimilar states s, t in which $m(s,t) = 0$, m satisfies $F(m)(s,t) \leq m(s,t) = 0$, For the non bisimilar states s, t in which $m(s,t) = \infty$, $F(m)(s,t) \leq m(s,t)$ holds. Henceforth, $m \preceq F(m)$, m is a pre-fixed-point of F.

Recall that m^K is the greatest pre-fixed-point of F, namely, $m^K = \bigsqcup\{m \in \mathcal{M} | m \preceq F(m)\}$. We have $m \preceq m^K$. $m^K(s,t) \leq m(s,t) = 0$, thus for the maximum fixed-point: $m^K(s,t) = 0$.

($\Leftarrow$) Consider the relation $\mathcal{R}$ induced by 0 distance in m^K. Clearly it is an equivalence relation. We now show that it is a weak bisimulation. Let $m^K(s,t) = 0$. Consider an arbitrary $\mathcal{R}$-closed set $[s_i] \in S/\mathcal{R}$, $\mu(s, a, [s_i]) = \sum_{s \in [s_i]} \mu_1(s) = \mu_1([s_i])$ for some $s \stackrel{a}{\Longrightarrow} \mu_1$. Since m^K is a fixed-point of F, there exists some μ_2 such that $t \stackrel{a}{\Longrightarrow} \mu_2$ and $m^K(\mu_1, \mu_2) = 0$. We will prove that $\mu_2([s_i]) \geq \mu_1([s_i])$.

We first show that

$$\sum_i \mu_1(s_i) = \sum_i \mu_2(s_i). \tag{20}$$

Consider the primal program (10), we know that $\ln \dfrac{\sum_i \mu_1(s_i)x_i}{\sum_i \mu_2(s_i)x_i}$ and $\ln \dfrac{\sum_i \mu_2(s_i)x_i}{\sum_i \mu_1(s_i)x_i}$ are bounded by 0. Let $x_i = 1$ for all $1 \leq i \leq |S|$, We obtain $\sum_i \mu_1(s_i) \leq \sum_i \mu_2(s_i)$ and $\sum_i \mu_1(s_i) \geq \sum_i \mu_2(s_i)$. Straightforwardly, it holds that $\sum_i \mu_1(s_i) = \sum_i \mu_2(s_i)$.

Using the dual programs, we know that the optimal values of z are not greater than 1. Consider the l_{ij}, r_i which achieve the minimum $\ln z$ in the dual program of $\ln \sum_i \mu_1(s_i)x_i - \ln \sum_i \mu_2(s_i)x_i$, where the optimal value $z \leq 1$. We shall prove that

$$\text{if } [s_i] \neq [s_j] \text{ then } l_{ij} = 0. \tag{21}$$

It satisfies that:

$$\mu_1(s_i) = \sum_j l_{ij} - r_i \tag{22}$$

$$z \cdot \mu_2(s_j) \geq \sum_i l_{ij} e^{m(s_i, s_j)} - r_j \tag{23}$$

Adding up all the constraint equations for μ_1 and μ_2 respectively, it follows that:

$$\begin{aligned}
\textstyle\sum_i \mu_1(s_i) &= \textstyle\sum_{i,j} l_{ij} - \sum_i r_i \\
z \cdot \textstyle\sum_j \mu_2(s_j) &\geq \textstyle\sum_{i,j} l_{ij} e^{m(s_i,s_j)} - \sum_j r_j \\
\textstyle\sum_{i,j} l_{ij} e^{m(s_i,s_j)} - \sum_j r_j &\leq \textstyle\sum_{i,j} l_{ij} - \sum_i r_i \quad \text{by equation (20) and } z \leq 1 \\
\textstyle\sum_{i,j} l_{ij}(e^{m(s_i,s_j)} - 1) &\leq 0 \\
l_{ij}(e^{m(s_i,s_j)} - 1) &= 0 \qquad \text{by } l_{ij} \geq 0 \text{ and } m^K(s_i, s_j) \geq 0
\end{aligned}$$

Thus, if $[s_i] \neq [s_j]$, i.e. $m^K(s_i, s_j) \neq 0$, then it must hold that $l_{ij} = 0$.

$$\begin{aligned}
\mu_2([s_i]) &= \textstyle\sum_{s_j \in [s_i]} \mu_2(s_j) \\
&\geq \textstyle\sum_{s_j \in [s_i]} \sum_k l_{kj} e^{m(s_k,s_j)} - \sum_{s_j \in [s_i]} r_j && \text{by (23)} \\
&= \textstyle\sum_{s_j \in [s_i]} \sum_{s_k \in [s_i]} l_{kj} e^{m(s_k,s_j)} - \sum_{s_j \in [s_i]} r_j && \text{by (21)} \\
&= \textstyle\sum_{s_j, s_k \in [s_i]} (l_{kj} - r_k) && \text{by } m^K(s_k, s_j) = 0 \\
&= \mu_1([s_i]) && \text{by (22) and (21)}
\end{aligned}$$

Hence $\mu_2([s_i]) \geq \mu_1([s_i])$ for all $[s_i] \in S/\mathcal{R}$, ensuring $\mu(t, a, [s_i]) \geq \mu_2([s_i]) \geq \mu_1([s_i]) = \mu(s, a, [s_i])$.

By the symmetry property of $\mathcal{R}$, we get $\mu(s, a, [s_i]) \geq \mu(t, a, [s_i])$ and therefore $\mu(s, a, [s_i]) = \mu(t, a, [s_i])$ as required. □

A.13 Proof of Proposition 3: the second item

$m^D(s, t) = 0 \Rightarrow s \approx t$. It is straightforwardly obtained from Lemma 7: $m^D \preceq m^K$ and the first item.

A.14 Proof of Proposition 3: the third item

$m^A(s, t) = 0 \Rightarrow s \approx t$.

Proof. Consider the relation $\mathcal{R}$ induced by 0 distance in m^A. Clearly it is an equivalence relation. We show that it is a weak bisimulation. Let $m^A(s, t) = 0$. Consider an arbitrary $\mathcal{R}$-closed set $[s_i] \in S/\mathcal{R}$, $\mu(s, a, [s_i]) = \sum_{s \in [s_i]} \mu_1(s) = \mu_1([s_i])$ for some $s \stackrel{a}{\Longrightarrow} \mu_1$. Since $m^A(s, t) = 0$, there exists an 0-amortised bisimulation $\mathcal{R}' \subseteq S \times S \times [0, 0]$ such that $(s, t, 0) \in \mathcal{R}'$. There exist a bijection β and a distribution μ_2 such that $t \stackrel{a}{\Longrightarrow} \mu_2$, for any $s_i \in \mathit{supp}(\mu_1)$, $t_i = \beta(s_i)$ and $(s_i, t_i, \ln \mu_1(s_i) - \ln \mu_2(t_i)) \in \mathcal{R}'$. Because the leakage budget is 0, which says that during the mutual simulation, every step must have exactly the same probability, i.e. $\mu_1(s_i) = \mu_2(t_i)$. Furthermore by $(s_i, t_i, 0) \in \mathcal{R}'$, we have $m^A(s_i, t_i) = 0$, thus $[s_i] = [t_i]$. Henceforth, $\mu_1([s_i]) = \sum_{s \in [s_i]} \mu_1(s) = \sum_{\beta(s) \in [s_i]} \mu_2(\beta(s)) = \mu_2([s_i])$ for all $[s_i] \in S/\mathcal{R}$, ensuring $\mu(t, a, [s_i]) \geq \mu_2([s_i]) = \mu_1([s_i]) = \mu(s, a, [s_i])$.

By the symmetry property of $\mathcal{R}$, we get $\mu(s, a, [s_i]) \geq \mu(t, a, [s_i])$ and therefore $\mu(s, a, [s_i]) = \mu(t, a, [s_i])$ as required. $\square$