1 DiPA with a Counter

1.1 DiPA* and Definitions

Definition 1.1. Fix parameters ϵ, N. Let C be the guard conditions $\{n < N, \texttt{true}, \texttt{insample} \geq \texttt{x}, \texttt{insample} < \texttt{x}, n < N \text{ AND } \texttt{insample} \geq \texttt{x}, n < N \text{ AND } \texttt{insample} < \texttt{x}, n \geq N\}$ A **DiP* automaton** (DiPA*) $\mathcal{A}$ is defined as the tuple $\mathcal{A} = (Q, \Sigma, \Gamma, q_0, X, P, \delta)$, where:

- Q = finite set of states; partitioned into input states Q_{in} and non-input states Q_{non}
- Σ is the input alphabet (taken to be $\mathbb{R}$)
- Γ is a finite output alphabet
- $q_0 \in Q$ is the starting state
- X= {x, insample, insample', n} is a set of variables. $\texttt{x}, \texttt{insample}, \texttt{insample'} \in \mathbb{R}$; $\texttt{n} \in \mathbb{N}$ and is initialized to 0.
- $P : Q \to \mathbb{Q}^{\geq 0} \times \mathbb{Q} \times \mathbb{Q}^{\geq 0} \times \mathbb{Q}$ describing the parameters for sampling from Laplace distributions at each state.
- $\delta : (Q \times C) \to Q \times (\Gamma \cup \{\texttt{insample}, \texttt{insample'}\} \cup \{\phi\}) \times \{\texttt{true}, \texttt{false}\} \times \{0, 1\}$ is the transition function (technically a relation) that defines what state to transition to, what symbol or real value to output, whether or not x is assigned to, and whether or not n is incremented based on the current state and transition guard.

There are certain conditions that δ must satisfy; these are almost all the same as the restrictions on transition functions of DiPA, but with some slight modifications and one major addition (marked in blue):

- **Determinism:** If $\delta(q, \texttt{true})$ is defined, then no other transitions out of q can be defined. Additionally, at most one of $\delta(q, \texttt{insample} \geq \texttt{x})$ and

  $\delta(q, n < N \text{ AND } \texttt{insample} \geq \texttt{x})$ can be defined and at most one of $\delta(q, \texttt{insample} < \texttt{x})$ and $\delta(q, n < N \text{ AND } \texttt{insample} < \texttt{x})$ can be defined.

  If $\delta(q, n < N)$ is defined, then $\delta(q, n < N \text{ AND } \texttt{insample} \geq \texttt{x})$ and $\delta(q, n < N \text{ AND } \texttt{insample} < \texttt{x})$ are not defined. Additionally, if any of $\delta(q, n < N)$, $\delta(q, n < N \text{ AND } \texttt{insample} < \texttt{x})$, or $\delta(q, n < N \text{ AND } \texttt{insample} \geq \texttt{x})$ are defined, then $\delta(q, n \geq N)$ must be defined as well. Finally, if $\delta(q, n \geq N)$ is defined, then $\delta(q, \texttt{true})$, $\delta(q, \texttt{insample} \geq \texttt{x})$ and $\delta(q, \texttt{insample} < \texttt{x})$ are not defined.

  For the sake of convenience, from now on, we will use $\texttt{true}$ to refer to both guards $\texttt{true}$ and $n < N$, $\texttt{insample} \geq \texttt{x}$ to refer to both $\texttt{insample} \geq \texttt{x}$ and $n < N$ AND $\texttt{insample} \geq \texttt{x}$, and $\texttt{insample} < \texttt{x}$ to refer to both $\texttt{insample} < \texttt{x}$ and $n < N$ AND $\texttt{insample} < \texttt{x}$.

- **Output Distinction:** For any state $q \in Q$, if $\delta(q, \texttt{insample} \geq \texttt{x}) = (q_1, o_1, b_1, i_1)$ and $\delta(q, \texttt{insample} < \texttt{x}) = (q_2, o_2, b_2, i_2)$, then $o_1 \neq o_2$ and at least one of $o_1 \in \Gamma$ and $o_2 \in \Gamma$ is true. In addition, $o_1 \neq \phi$ and $o_2 \neq \phi$ and if $\delta(q, n \geq N) = (q', o', b', i')$, then $o' = \phi$,

i.e., the ϕ output symbol is reserved for transitions with guard $n \geq N$, which must output ϕ.

- **Initialization:** The initial state q_0 has only one outgoing transition of the form $\delta(q_0, \texttt{true}) = (q, o, \texttt{true}, i)$ for $i \in \{0, 1\}$.
- **Non-input transition:** From any $q \in Q_{non}$, if $\delta(q, c)$ is defined, then $c = \texttt{true}$.
- **Control Flow Separation:** Consider the underlying graph G of $\mathcal{A}$. For all states $q \in Q$, if $\delta(q, n \geq N) = (q', o, b, i)$,$q$ and q' must be in different strongly connected components of G.

Note that the **control flow separation** condition implies that no cycle in G can contain an edge that corresponds to a transition with guard $n \geq N$. In addition, determinism combined with control flow separation imply that no two transitions (i.e. transitions with different guards) can be from some state q to the same state q'.

1.1.1 Path Probabilities

Definition 1.2. (summarized from [1]) A **path** ρ of length n of a DiPA* $\mathcal{A}$ is a sequence of states, inputs, and outputs $\rho = q_0 \xrightarrow{a_0, o_0} q_1 \to \cdots \to q_{n-1}$, where q_i are the states traversed in $\mathcal{A}$, a_i are the inputs read in each state q_i, and o_i are the outputs output by $\mathcal{A}$ at the transition $q_i \to q_{i+1}$. We denote the sequence of inputs a_i for a path ρ as $\texttt{inseq}(\rho)$ and the sequence of outputs o_i as $\texttt{outseq}(\rho)$. In general, for a path $\rho = q_0 \to q_1 \to \cdots \to q_{n-1}$ we denote the transition $q_i \to q_{i+1}$ by $\texttt{trans}(q[i])$ and the guard of $\texttt{trans}(\rho[i])$ as $\texttt{guard}(\rho[i])$.

Definition 1.3. (from [1]) Two paths $\rho = q_0 \xrightarrow{a_0, o_0} q_1 \to \cdots \to q_n$ and $\rho' = q'_0 \xrightarrow{a'_0, o'_0} q'_1 \to \cdots \to q'_n$ of a DiPA* $\mathcal{A}$ are **equivalent** if for all i, $o_i = o'_i$ and $q_i = o'_i$. In other words, ρ and ρ' traverse the same states in $\mathcal{A}$ and produce the same output, and only possibly differ in the inputs they read.

For any path ρ of a DiPA* $\mathcal{A}$, we define $\mathbb{P}[\epsilon, N, x, n, \rho]$ as the **probability** of path ρ being traversed with $\mathcal{A}$ parameters ϵ and N, stored value x, and counter value n. This will enable us to define what it means for a DiPA* to be differentially private.

Consider a path $\rho = q_0 \xrightarrow{a_0, o_0} q_1 \to \cdots \xrightarrow{a_{n-1}, o_{n-1}} q_n$. Here, a_i and o_i are the input to state q_i and output of transition $q_i \to q_{i+1}$, respectively (if q_i does not take in input, $a_i = 0$).

If $|\rho| = 0$,, we define $\mathbb{P}[\epsilon, N, x, n, \rho] = 1$. Otherwise, we define $\mathbb{P}[\epsilon, N, x, n, \rho]$ recursively: Let $P(q_0) = (d, \mu, d', \mu')$ be the parameters for sampling from Laplace distributions for $\texttt{insample}$ and $\texttt{insample}'$. Let (q_0, c, q_1, o_0, b, i) represent the 0th transition, where c is the guard of the 0th transition, b is whether or not the 0th transition is an assignment transition, and i is the amount that the counter n gets incremented by in the 0th transition.

Let $\nu = \mu + a_0$. If $o_0 = (y, v, w)$ for $y \in \{\texttt{insample}, \texttt{insample'}\}$, then let

$$k = \int_v^w \frac{d\epsilon}{2} e^{-d\epsilon|z-\mu-a_0|} dz$$
$$k' = \int_v^w \frac{d'\epsilon}{2} e^{-d'\epsilon|z-\mu'-a_0|} dz$$

If the 0th transition of ρ is not an assignment transition (i.e. $b = \texttt{false}$), then we define $\mathbb{P}[\epsilon, N, x, n, \rho]$ as follows:

Case 1: $n \geq N$ and $c = n \geq N$. If $o_0 \in \Gamma$, then $\mathbb{P}[\epsilon, N, x, n, \rho] = \mathbb{P}[\epsilon, N, x, n+i, \texttt{tail}(\rho)]$. If $o_0 = (\texttt{insample}, v, w)$ then $\mathbb{P}[\epsilon, N, x, n, \rho] = k\mathbb{P}[\epsilon, N, x+i, \texttt{tail}(\rho)]$. If $o_0 = (\texttt{insample}, v, w)$ then $\mathbb{P}[\epsilon, N, x, n, \rho] = k'\mathbb{P}[\epsilon, N, x, n+i, \texttt{tail}(\rho)]$

Case 2: $n < N$ and $c = n \geq N$. Then we define $\mathbb{P}[\epsilon, N, x, n, \rho] = 0$.

Every case for other guards is exactly analogous to their counterpart definitions in [1], but in general where $\mathbb{P}[\epsilon, N, x, n, \texttt{tail}(\rho)]$ is referenced in [1], $\mathbb{P}[\epsilon, N, x, n+i, \texttt{tail}(\rho)]$ should be used instead.

Because of the initialization condition, for paths starting at the start state of $\mathcal{A}$, the starting value of x is irrelevant. In addition, since n is always initialized to 0, we will abuse notation for paths ρ that start at the start state to write $\mathbb{P}[\epsilon, N, \rho]$ to represent $\mathbb{P}[\epsilon, N, x, 0, \rho]$.

We can use this definition of path probabilities to formalize what it means for paths to be valid program traces in $\mathcal{A}$:

Definition 1.4. A path $\rho = q_0 \to q_1 \to \cdots q_n$ from the start state q_0 of $\mathcal{A}$ is **valid** if $\mathbb{P}[\epsilon, N, \rho] > 0$.

Most notably, given a definition of path probabilities, we can define what it means for a DiPA* to be differentially private:

Definition 1.5. As in [1], a DiPA* $\mathcal{A}$ with parameters ϵ, N is $d\epsilon-$**differentially private** if for all equivalent paths ρ, ρ' in $\mathcal{A}$ such that $\texttt{inseq}(\rho)$ and $\texttt{inseq}(\rho')$ are adjacent, $\mathbb{P}[\epsilon, N, \rho] \leq e^{d\epsilon}\mathbb{P}[\epsilon, N, \rho']$.

1.2 Violations of Differential Privacy

NOTE: this section is obsolete now, remove

Definition 1.6. An **abstract path** in a DiPA* is a sequence of states and associated outputs $q_0\sigma_0 q_1\sigma_1 \cdots q_{n-1}\sigma_{n-1}q_n$.

Abstract paths will be used almost entirely for the purposes of analysis.

Definition 1.7. A **bounded** cycle C in a DiPA* $\mathcal{A}$ is a cycle in $\mathcal{A}$ where there exists at least one transition $(q', \sigma, t, 1)$ (i.e. `n` gets incremented) and there exists some $q \in Q$ ("exit state") in the cycle such that $f(q, n \geq N) = (q', \sigma, t, i)$ where q' is not in the cycle. Otherwise, the cycle is **unbounded**.

Definition 1.8. A cycle C with an exit state with transition $n \geq N$ is an **infeasible** cycle if, for *all* paths $\rho = q_0 \to q_1 \to \cdots \to q_m$ from the start state to a state $q_m \in C$, at least N transitions $q_i \to q_{i+1}$ are increment transitions or some transition $q_i \to q_{i+1}$ has guard $n \geq N$. Otherwise, C is **feasible**.

Definition 1.9. (from [1]) A **leaking cycle** is a cycle $C = q_0 \xrightarrow{a_0,o_0} q_1 \to \cdots \to q_{n-1} \to q_0$ in a DiPA* $\mathcal{A}$ if there exist indices $0 \leq i < j < n$ such that the ith transition $q_i \to q_{i+1}$ is an assignment transition and the guard of the jth transition guard is not $n < N$.

Proposition 1.10. *If a DiPA* $\mathcal{A}$ has a reachable feasible unbounded leaking cycle, then it is not differentially private.*

Proof. Let $C = a_1 a_2 \cdots a_{n-1} a_n; a_1 = a_n$ be such a cycle in $\mathcal{A}$. We will reduce the analysis to an analogous DiPA.

Case 1: C does not have an exit state.

Consider an abstract path $\eta = q_0 \sigma_0 q_1 \cdots q_{m+n-1} \sigma_{m+n-1} q_m$ such that $a_1 \cdots a_n = q_m \cdots q_{m+n}$ (i.e. the last n states of the path are the cycle C).

For $\ell > 0$, let η_ℓ be the abstract path $\eta_\ell = q_0 \sigma_0 q_1 \sigma_1 \cdots q_{m+\ell n-1} \sigma_{m+\ell n-1} q_{m+\ell n}$ such that $q_k = q_{k-n}$ and $\sigma_k = \sigma_{k-n}$ for all $m + n \leq k \leq m + \ell n$. This is the path η with the cycle C repeated ℓ times. Note that because C has no exit state, for all states $a_i \in C$, all transitions from a_i have a guard that is *not* $n \geq N$. This means that the path η_ℓ in $\mathcal{A}$ exists for all $\ell > 0$. Thus, the same input sequences α_ℓ and β_ℓ as described in Lemma 6 of [1] are witnesses to a violation of differential privacy. In particular, the same analysis holds because there is some fixed number f such that η_ℓ has at most f transitions with guard $n \geq N$, even as ℓ can vary arbitrarily.

Case 2: Suppose that C has no increment transition.

Because C is feasible, there exists some abstract path $\eta = q_0 \sigma_0 q_1 \cdots q_{m+n-1} \sigma_{m+n-1} q_m$ such that $a_1 \cdots a_m = q_m \cdots q_{m+n}$ and at $q_m = a_1$, $n < N$.

As in Case 1, for $\ell > 0$, consider $\eta_\ell = q_0 \sigma_0 q_1 \sigma_1 \cdots q_{m+\ell n-1} \sigma_{m+\ell n-1} q_{m+\ell n}$ such that $q_k = q_{k-n}$ and $\sigma_k = \sigma_{k-n}$ for all $m + n \leq k \leq m + \ell n$. Because there are no increment transitions in C, $\forall 0 \leq i \leq \ell n$, `true` at state q_i. So for all states $a_i \in C$, a transition from a_i with guard $n \geq N$ will never be taken by $\mathcal{A}$. As before, then, the path η_ℓ in $\mathcal{A}$ exists for all $\ell > 0$, so α_ℓ and β_ℓ from Lemma 6 of [1] are witnesses to a violation of differential privacy.

□

Definition 1.11. (from [1]) A cycle ρ of a DiPA* $\mathcal{A}$ is an `L`-cycle (respectively, `G`-cycle) if there is an $i < |\rho|$ such that $\texttt{guard}(\rho[i]) = \texttt{insample} < \texttt{x}$ (respectively $\texttt{guard}(\rho[i]) = \texttt{insample} \geq \texttt{x}$).

Definition 1.12. (from [1]) A path ρ of a DiPA $\mathcal{A}^*$ is an `AL`-path (respectively, `AG`-path) if all assignment transitions on ρ have guard $\texttt{insample} < \texttt{x}$ (respectively, $\texttt{insample} \geq \texttt{x}$).

Definition 1.13. (from [1]) A pair of cycles (C, C') in a DiPA* $\mathcal{A}$ is a **leaking pair** if one of the following is satisfied:

- C is an `L`-cycle, C' is a `G`-cycle, and there is an `AG`-path from a state in C to a state in C'.
- C is an `G`-cycle, C' is a `L`-cycle, and there is an `AL`-path from a state in C to a state in C'.

Definition 1.14. A pair of cycles (C, C') is a feasible unbounded leaking pair of cycles if both C and C' are feasible and unbounded cycles, C is an `L`-cycle (respectively, `G`-cycle), C' is a `G`-cycle (respectively `L`-cycle), and there exists an `AL`-path (respectively, `AG`-path) $\rho = a_1 a_2 \cdots a_k$ from C to C' (i.e. such that $a_1 \in C$ and $a_k \in C'$) such that all of the following hold:

1. Either there are no $n \geq N$ transitions on ρ or C' has no exit state.
2. Either there exists some path τ from the start state q_0 of $\mathcal{A}$ to a_k that includes a_1 such that there are at most $N - 1$ increment transitions on τ or C' has no exit state.
3. Either C' has no exit state or C has no increment transitions.
4. If there exists an $n \geq N$ transition in ρ from states a_i to a_{i+1}, there exists some path τ from the start state q_0 of $\mathcal{A}$ to a_i that includes a_1 such that there are at least N increment transitions in τ.

Conditions (1)-(3) ensure that there exist some path in $\mathcal{A}$ such that either $n < N$ when entering C' or that C' has no exit state; otherwise, C' would be rendered infeasible in practice.

Condition (4) ensures that the path ρ between C and C' is in fact traversible.

Proposition 1.15. *If a DiPA* $\mathcal{A}$ has a feasible unbounded leaking pair of cycles (C, C') where C is reachable, then it is not differentially private.*

Proof. As with proposition 2.4, we reduce this case to a similar case in a DiPA.

Because of proposition 2.4, we can assume that $\mathcal{A}$ does not have a leaking cycle. Assume that $\mathcal{A}$ has a feasible unbounded pair of leaking cycles (C, C') such that C is reachable from the start state of $\mathcal{A}$. Suppose that C is an `L`-cycle and C' is a `G`-cycle. Let $\rho = a_1 \cdots a_m$ be an `AL`-path from C to C'. The proof of the symmetric case is analogous.

Because we assume that there are no leaking cycles, there are no assignment transitions in C and C'. WLOG assume that C and C' are distinct, and finally assume that all states in $\mathcal{A}$ are input states. Let $n_1 = |C|, n_2 = |C'|$.

We want to construct a valid abstract path

$$\eta_\ell = q_0 \sigma_0 \cdots q_u \sigma_u \cdots q_\nu \sigma_\nu \cdots q_{v+n_1\ell-1} \sigma_{v+n_1\ell-1} \cdots q_w \sigma_w \cdots o_{w+n_2\ell-1} q_{w+n_2\ell}$$

for all $\ell > 0$. Let t_k be the k-th transition of η and c_k be the guard of t_k. ν_ℓ must satisfy the following conditions:

1. q_0 is the start state of $\mathcal{A}$
2. $q_v\sigma_v q_{v+1}\sigma_{v+1}\cdots q_{v+n_1-1}\sigma_{v+n_1-1}q_{v+n_1}$ is the cycle C
3. $t_{j+n_1} = t_j$ for all $j, v \leq j < v + n_1(\ell - 1)$
4. $q_w\sigma_w q_{w+1}\sigma_{w+1}\cdots q_{w+n_2-1}\sigma_{w+n_2-1}q_{w+n_2}$ is the cycle C'
5. $t_{j+n_2} = t_j$ for all $j, w \leq j < w + n_2(\ell - 1)$
6. t_u is an assignment transition and $\forall j, u < j < v + n_1\ell$ and $\forall j, j \geq w$, t_j is a non-assignment transition.
7. For all $j, v + n_1\ell \leq j < w$, if t_j is an assignment transition then $c_j =$ `insample` $\geq$ `x`

Intuitively, $\forall \ell > 0, \eta_\ell$ is a path that begins at the start state, reaches C, traverses C ℓ times, then traverses an `AG`$-$path to C' before traversing C' ℓ times. If such a path exists in $\mathcal{A}$, then as in proposition 2.5, the adjacent input sequences $\alpha(\ell)$ and $\beta(\ell)$ from lemma 7 of [1] serve as witnesses to the violation of differential privacy.

Since C' is unbounded, C' must either lack an exit state or an increment transition (or both).

Let γ and γ' be words representing the states in the cycles C and C', respectively.

Case 1: C' does not have an exit state.

Since C is reachable, there is a valid path from the start state q_0 to a state c_1 in C. Let s be the word representing states on such a path from q_0 to c_1.

Since C is feasible and unbounded, the path represented by the word $s \cdot \gamma^\ell$ is valid.

Similarly, if there are no transitions in ρ that have guard $n \geq N$, then $s \cdot \gamma^\ell \cdot \rho$ is valid for all $\ell > 0$.

If there exists a transition ρ, then by condition (4) from the definition of feasible unbounded leaking pairs, for all such transitions $a_i \to a_{i+1}$, there exists some path $\tau = s' \cdot \rho'$ from q_0 to a_i such that s' is a valid path from q_0 to a_1 and ρ' is a subpath of ρ from a_1 to a_i. Since $a_1 \in C$, this means that the path represented by the word $s' \cdot \gamma^\ell \cdot \rho$ is still valid for all $\ell > 0$.

In either case, there exists some word $\alpha \cdot \gamma^\ell \cdot \rho$ that represents a valid path from the start state.

Since C' has no exit state, no transition in C' can have guard $n \geq N$. To see this, consider a state $q \in C'$ that has a transition with guard $n \geq N$ to another state q'. If $q' \notin C'$, then q is an exit state. However, if $q' \in C'$, then control flow separation is violated. Therefore, all transitions in C' must be one of `true`, `insample` $\geq$ `x`, or `insample` $<$ `x`, which means that a path can traverse C' ℓ times for any $\ell > 0$ with non-zero probability.

Thus, the path represented by the word $\eta_\ell = \alpha \cdot \gamma^\ell \cdot \rho \cdot \gamma^\ell$ is valid for all $\ell > 0$.

Case 2: C' has an exit state, but no increment transition.

By condition (2) of the definition of feasible unbounded leaking pairs, we know that there exists some path $\tau = \alpha \cdot \rho$ where α is a path from the start state to a_1 such that there are no more than $N-1$ increment transitions in τ. Notably, this path is valid because C is reachable and because ρ does not have any transitions with guard $n \geq N$ by condition (1). Since $a_1 \in C$, as before, the path represented by the word $\alpha \cdot \gamma^\ell \cdot \rho$ is valid for all $\ell > 0$. In addition, such a path still only has at most $N-1$ increment transitions, since there are no increment transitions in C as per condition (3).

Due to control flow separation, we can note that every transition in C' has a guard of $\texttt{true}, \texttt{insample} \geq \texttt{x}$, or $\texttt{insample} < \texttt{x}$. In addition, because when following such a path τ, $n < N$ at $a_k \in C'$, and C' has no increment transitions, we can loop through C' an arbitrary number of times without any $n \geq N$ guards being satisfied.

Thus, the path represented by the word $\eta_\ell = \alpha \cdot \gamma^\ell \cdot \rho \cdot \gamma'^\ell$ is valid for all $\ell > 0$. □

Definition 1.16. (from [1]) A cycle C of a DiPA $\mathcal{A}$ is a **disclosing cycle** if there exists some $0 \leq i < |C|$ such that $\texttt{trans}(C[i])$ is an input transition that outputs either $\texttt{insample}$ or $\texttt{insample}'$.

Proposition 1.17. *If a DiPA* $\mathcal{A}$ has a reachable feasible unbounded disclosing cycle, then it is not differentially private.*

The proof of proposition 2.12 is almost identical to the proof of proposition 2.4:

Proof. Let $C = a_1 a_2 \cdots a_{n-1} a_n; a_1 = a_n$ be such a cycle in $\mathcal{A}$. We will reduce the analysis to an analogous DiPA.

Case 1: C does not have an exit state.

Consider an abstract path $\eta = q_0 \sigma_0 q_1 \cdots q_{m+n-1} \sigma_{m+n-1} q_m$ such that $a_1 \cdots a_n = q_m \cdots q_{m+n}$ (i.e. the last n states of the path are the cycle C).

For $\ell > 0$, let η_ℓ be the abstract path $\eta_\ell = q_0 \sigma_0 q_1 \sigma_1 \cdots q_{m+\ell n-1} \sigma_{m+\ell n-1} q_{m+\ell n}$ such that $q_k = q_{k-n}$ and $\sigma_k = \sigma_{k-n}$ for all $m+n \leq k \leq m+\ell n$. This is the path η with the cycle C repeated ℓ times. Note that because C has no exit state, for all states $a_i \in C$, all transitions from a_i have a guard that is *not* $n \geq N$. This means that the path η_ℓ in $\mathcal{A}$ exists for all $\ell > 0$. Thus, the same input sequences α_ℓ and β_ℓ as described in Lemma 8 of [1] are witnesses to a violation of differential privacy. In particular, the same analysis holds because there is some fixed number f such that η_ℓ has at most f transitions with guard $n \geq N$, even as ℓ can vary arbitrarily.

Case 2: Suppose that C has no increment transition.

Because C is feasible, there exists some abstract path $\eta = q_0 \sigma_0 q_1 \cdots q_{m+n-1} \sigma_{m+n-1} q_m$ such that $a_1 \cdots a_m = q_m \cdots q_{m+n}$ and at $q_m = a_1$, $n < N$.

As in Case 1, for $\ell > 0$, consider $\eta_\ell = q_0 \sigma_0 q_1 \sigma_1 \cdots q_{m+\ell n-1} \sigma_{m+\ell n-1} q_{m+\ell n}$ such that $q_k = q_{k-n}$ and $\sigma_k = \sigma_{k-n}$ for all $m+n \leq k \leq m+\ell n$. Because there are no increment transitions in C, $\forall 0 \leq i \leq \ell n$, $\texttt{true}$ at state q_i. So for all states $a_i \in C$, a transition from a_i with guard

$n \geq N$ will never be taken by $\mathcal{A}$. As before, then, the path η_ℓ in $\mathcal{A}$ exists for all $\ell > 0$, so α_ℓ and β_ℓ from Lemma 8 of [1] are witnesses to a violation of differential privacy. □

Definition 1.18. (adapted from [1]) An feasible unbounded privacy violating lasso is a path $\rho = a_1 \to a_2 \to \cdots \to a_k$ of length n in a DiPA* $\mathcal{A}$ such that one of the following hold:

- $\texttt{tail}(\rho)$ is an AG-path (respectively, AL-path) such that $\texttt{last}(\rho)$ is in a feasible unbounded G-cycle (respectively, L-cycle) and the 0th transition is an assignment transition that outputs `insample`.
- ρ is an AG-path (respectively, AL-path) such that $\texttt{first}(\rho)$ is in a feasible unbounded G-cycle (respectively, L-cycle) and the 0th transition has guard $\texttt{insample} < \texttt{x}$ (respectively, $\texttt{insample} \geq \texttt{x}$) and outputs `insample`
- ρ is an AG-path (respectively, AL-path) such that $\texttt{first}(\rho)$ is in a feasible unbounded L-cycle (respectively, G-cycle) and the last transition has guard $\texttt{insample} \geq \texttt{x}$ (respectively, $\texttt{insample} < \texttt{x}$) and outputs `insample`.

In addition, if there are any transitions $a_i \to a_{i+1}$ in ρ with guard $n \geq N$, there must exist some path represented by the word $\tau = \alpha \cdot \beta$ from the start state of $\mathcal{A}$ to a_i such that α represents a path from the start state of $\mathcal{A}$ to a_1 and β represents a subpath of ρ from a_1 to a_i.

Definition 1.19. For a lasso ρ, let C_ρ be the cycle associated with[1] ρ. Then a lasso ρ in a DiPA* $\mathcal{A}$ is bounded iff C_ρ is bounded. Similarly, ρ is feasible iff C_ρ is feasible.

Proposition 1.20. *If a DiPA* $\mathcal{A}$ has a reachable unbounded feasible privacy violating lasso, then it is not differentially private.*

Proof. As in [1], we will only prove the third case of privacy violating paths here.

Let $\rho = a_1 \cdots a_k$ be an AG-path such that $\texttt{first}(\rho) = a_1$ is in an L-cycle C where the last transition of ρ has guard $\texttt{insample} \geq \texttt{x}$ and outputs `insample`.

Similar to propositions 2.4, 2.10, and 2.12, we will construct a valid path η_ℓ in $\mathcal{A}$ starting at the start state that traverses C ℓ times and then traverses the path ρ for arbitrary $\ell > 0$. With this path, the input sequences $\alpha(\ell)$ and $\beta(\ell)$ from lemma 9 of [1] serve as witnesses for a violation of differential privacy.

Let γ be the word representing C.

If there are any transitions in ρ with guard $n \geq N$, then from the definition of feasible unbounded privacy violating lassos, we know that there exists some path α from the start state to a_1 such that all such guards will be satisfied on ρ.

Case 1: C does not have an exit state.

As before, this means that $\alpha \cdot \gamma^\ell$ is a valid path for all $\ell > 0$, and thus that $\alpha \cdot \gamma^\ell \cdot \rho$ is valid for all $\ell > 0$.

[1] Hopefully this is clear

Case 2: C has no increment transition. Since C is feasible, this also means that $\alpha \cdot \gamma^\ell$ is a valid path for all $\ell > 0$, and thus that $\alpha \cdot \gamma^\ell \cdot \rho$ is valid for all $\ell > 0$.

□

Definition 1.21. A DiPA* $\mathcal{A}$ is well-formed if $\mathcal{A}$ has no reachable unbounded feasible leaking cycles, unbounded feasible leaking pair (C, C') where C is reachable, reachable unbounded feasible disclosing cycles, or reachable unbounded feasible privacy violating lassos.

Theorem 1.22. *If a DiPA* is not well-formed, then it is not differentially private.*

Proof. Follows from propositions 2.4, 2.10, 2.12, and 2.15. □

2 Proving Differential Privacy

Theorem 2.1. *A DiPA* is well-formed iff it is differentially private.*

Proof. Let $\mathcal{A}^* = (Q, \Sigma, \Gamma, q_0, X^*, P^*, \delta^*)$ be a well-formed DiPA* with parameters ϵ and N.

Let $G = \{\texttt{true}, n < N, n \geq N, \texttt{insample} \geq \texttt{x}, n < N \text{ AND } \texttt{insample} \geq \texttt{x}, \texttt{insample} < \texttt{x}, n < N \text{ AND } \texttt{insample} < \texttt{x}\}$ be the set of guard conditions for DiPA*s.

Construct the DiPA $\mathcal{A} = (Q \times [N], \Sigma, \Gamma \cup \{\phi\}, (q_0, 0), X, P, \delta)$ as follows:

For each state $q \in Q^*$:

For $g \in G$, if $\delta^*(q, g) = (q', \sigma, \mathbf{b}, x)$ is defined, define $\delta((q, k), g)$ as follows:

Case 1: $g \in \{\texttt{true}, \texttt{insample} \geq \texttt{x}, \texttt{insample} < \texttt{x}\}$

For all $k \in [N-1]$, define the transitions

$$\delta((q, k), g) = ((q', k + x), \sigma, \mathbf{b})$$

and define the transition

$$\delta((q, N), g) = ((q', N), \sigma, \mathbf{b})$$

Case 2: $g = n \geq N$

We define the transition

$$\delta((q, N), \texttt{true}) = ((q', N), \sigma, \mathbf{b})$$

Case 3: $g = n < N$

For all $k \in [N-1]$, define the transitions

$$\delta((q, k), \texttt{true}) = ((q', k + x), \sigma, \mathbf{b})$$

Case 4: $g = n < N$ **AND** $\texttt{insample} < \texttt{x}$

For all $k \in [N-1]$, define the transitions

$$\delta((q,k), \texttt{insample} < \texttt{x}) = ((q', k+x), \sigma, \mathbf{b})$$

Case 5: $g = n < N$ **AND** $\texttt{insample} \geq \texttt{x}$

For all $k \in [N-1]$, define the transitions

$$\delta((q,k), \texttt{insample} \geq \texttt{x}) = ((q', k+x), \sigma, \mathbf{b})$$

Intuitively, at state (q,k) in $\mathcal{A}$, k will track the value of n in $\mathcal{A}^*$ (since everything above N is treated the same, we compress all of those values together).

For each state $(q,k) \in Q$, let $P((q,k)) = P^*(q)$.

Claim 2.2. $\mathcal{A}$ is a valid DiPA.

Proof. To be a valid DiPA, δ must satisfy four conditions: determinism, output distinction, initialization, and non-input transition.

Determinism: Consider some state $(q,k) \in Q$ and suppose that $\delta((q,k), \texttt{true})$ is defined. Then either $\delta^*(q, \texttt{true})$ is defined, $\delta^*(q, n < N)$ and $k < N$, or $\delta^*(q, n \geq N)$ is defined and $k = N$. By the condition of determinism for DiPA*s, if $\delta^*(q, \texttt{true})$, then no other transitions from q in $\mathcal{A}^*$ are defined. Thus, neither $\delta((q,k), \texttt{insample} < \texttt{x})$ or $\delta((q,k), \texttt{insample} \geq \texttt{x})$ can be defined in $\mathcal{A}$.

If $\delta^*(q, n < N)$ is defined and $k < N$, then similarly none of $\delta^*(q, n < N$ AND $\texttt{insample} \geq \texttt{x})$, $\delta^*(q, n < N$ AND $\texttt{insample} < \texttt{x})$, $\delta^*(q, \texttt{true})$, $\delta^*(q, \texttt{insample} < \texttt{x})$, or $\delta^*(q, \texttt{insample} \geq \texttt{x})$ can be defined. Additionally, since $n < N$, there is no additional transition from q corresponding to an $n \geq N$ guard in $\mathcal{A}^*$, so neither $\delta((q,k), \texttt{insample} < \texttt{x})$ or $\delta((q,k), \texttt{insample} \geq \texttt{x})$ can be defined in $\mathcal{A}$.

Similarly, if $\delta^*(q, n \geq N)$ is defined and $k = N$, no other transitions from q can be defined in $\mathcal{A}$.

Output distinction: This follows immediately from the output distinction condition of DiPA*s.

Initialization: By the initialization condition of DiPA*s, the initial state q_0 has only one outgoing transition of the form $\delta(q_0, \texttt{true}) = (q, o, \texttt{true}, i)$ for $i \in \{0, 1\}$. Thus, there is only one transition out of $(q_0, 0)$ in $\mathcal{A}$, with guard $\texttt{true}$.

Non-input transition: Since input transitions are preserved from $\mathcal{A}^*$ in the construction of $\mathcal{A}$, this follows immediately from the condition of non-input transition for DiPA*s. □

Lemma 2.3. *If $\mathcal{A}^*$ is well-formed if and only if $\mathcal{A}$ is well-formed.*

Proof. We will prove this using the following lemma:

Lemma 2.4. *If there exists a reachable cycle* $C = (a_0, k_0) \to (a_1, k_1) \to \cdots \to (a_{m-1}, k_{m-1}) \to (a_0, k_0)$ *in* $\mathcal{A}$ *if and only if there exists a reachable unbounded feasible cycle* $C^* = a_0 \to a_1 \to \cdots a_{m-1} \to a_0$ *in* $\mathcal{A}^*$.

Proof. Let $C = (a_0, k_0)(a_1, k_1) \cdots (a_{m-1}, k_{m-1})(a_0, k_0)$ be a cycle in $\mathcal{A}$. Note that by construction, there must exist a cycle $C^* = a_0 a_1 \cdots a_{m-1} a_0$ in $\mathcal{A}^*$.

Additionally note that for any path $(q, k) \to (q', k')$ in $\mathcal{A}$, $k' \geq k$. This implies that $k_0 = k_1 = \cdots = k_{m-1}$.

In order for C^* to be bounded, there must be an increment transition between some states a_i and a_{i+1} in C^*. However, this would mean that there exists a transition $(a_i, k_i) \to (a_{i+1}, k_i + 1)$, which is impossible because all k_i's are equal for $0 \leq i \leq m-1$ and because $\mathcal{A}$ is deterministic. Therefore C^* is unbounded.

Consider the underlying graphs $G_{\mathcal{A}}, G_{\mathcal{A}^*}$ of $\mathcal{A}$ and $\mathcal{A}^*$, respectively. Consider some edge $e^* = (q, q') \in G_{\mathcal{A}^*}$; there exists exactly one corresponding edge $e = ((q, k), (q', k'))$ in $G_{\mathcal{A}}$.

Suppose that C^* is infeasible for the sake of contradiction. Then $k_0 = k_1 = \cdots k_{m-1} = n$ and there is some exit state $s \in C^*$ such that $\delta^*(s, n \geq N) = (s', \sigma, \mathbf{b}, x)$. Note that because of determinism, there is exactly one transition $(s, n) \to (s', n)$ in G out of the state (s, n). Therefore, (s', n) must also be in the cycle C, which implies s' must be in the cycle C^*. However, this contradicts control flow separation, since then s and s' would be in the same component of $G_{\mathcal{A}^*}$, even with the edge corresponding to the $n \geq N$ transition removed. Thus, C^* is feasible.

Now let $C^* = a_0 \to \ldots \to a_m \to a_0$ be an reachable unbounded feasible cycle in $\mathcal{A}^*$. Because C^* is feasible, there exists some path from the start state q_0 of $\mathcal{A}^*$ to a_0 such that $n < N$ at a_0. Let k_0 be the minimum such value of n at a_0. Then consider the sequence of states $(a_0, k_0), (a_1, k_1), \ldots, (a_m, n_m), (a_0, k_0)$ in $\mathcal{A}$ such that k_i is defined as follows:

Let $(a_{i-1}, c, a_i, o, b, j)$ represent the transition from a_{i-1} to a_i in $\mathcal{A}^*$. Then for $i > 0, k_i = k_{i-1} + j$. Note that for all $1 \leq i \leq m$, there exists a transition $(a_{i-1}, k_{i-1}) \to (a_i, k_i)$ by construction.

Suppose that there is no increment transition in C^*. Then $\forall 1 \leq i \leq m, k_i = k_{i-1}$. Thus, $k_m = k_0$ and there exists a transition $(a_m, k_0) \to (a_0, k_0)$ in $\mathcal{A}$. Thus, the cycle $C = (a_0, k_0) \to (a_1, k_1) \to \ldots \to (a_m, k_m) \to (a_0, k_0)$ is a cycle in $\mathcal{A}$.

Now suppose that C^* has an increment transition but no exit state. Then the state (a_0, N) is reachable in $\mathcal{A}$. To see this, consider the path $(a_0, k_0) \to (a_1, k_1) \to \ldots \to (a_m, k_m) \to (a_0, k_0')$ in $\mathcal{A}$. Since C^* has an increment transition, $k_0' > k_0$. Since C^* has no exit state and k_i is bounded above by N, this means that there exists a path from $(q_0, 0)$ to (a_0, k_0) to (a_0, N). Then again because C^* has no exit state, $C = (a_0, N) \to (a_1, N) \to \ldots \to (a_m, N) \to (a_0, N)$ is a cycle in $\mathcal{A}$. □

Additionally, observe that if $(q, k) \to (q', k')$ is an assignment transition in $\mathcal{A}$, $q \to q'$ is also an assignment transition in $\mathcal{A}^*$. Similarly, if a transition $(q, k) \to (q', k')$ has a guard of `insample` $\geq$ `x` (respectively, `insample` $<$ `x`) in $\mathcal{A}$, $q \to q'$ also has a guard of `insample` $\geq$ `x`

(respectively, $\texttt{insample} < \texttt{x}$) in $\mathcal{A}^*$. Together, these mean that leaking cycles, leaking pairs, disclosing cycles, and privacy violating lassos in $\mathcal{A}$ correspond to their feasible unbounded equivalents in $\mathcal{A}^*$.

□

Lemma 2.5. *Let $\Psi = \{\rho : \rho \text{ is a path in } \mathcal{A}\}$ and $\Psi^* = \{\rho^* : \rho^* \text{ is a valid path in } \mathcal{A}^*\}$ be the sets of paths in $\mathcal{A}$ and $\mathcal{A}^*$, respectively. There exists a bijection $f : \Psi \to \Psi^* \times [N]$ such that $\forall x, \forall \rho \in \Psi$, if $f(\rho) = (\rho^*, n), \mathbb{P}[\epsilon, x, \rho] = \mathbb{P}[\epsilon, N, x, n, \rho^*]$.*

Proof. Let $\rho = (q_1, n_1) \to (q_2, n_2) \to \ldots \to (q_m, n_m)$ be a path in $\mathcal{A}$. Then let $f(\rho) = (q_1 \to q_2 \ldots \to q_m, n_1)$ such that $\texttt{inseq}(q_1 \to q_2 \ldots \to q_m) = \texttt{inseq}(\rho)$. Note that $\texttt{outseq}(q_1 \to q_2 \ldots \to q_m) = \texttt{outseq}(\rho)$ by output determinism.

By construction[2], $\rho^* = q_1 \to \ldots \to q_m$ must be a valid path in $\mathcal{A}^*$ if the value of the variable n in $\mathcal{A}^*$ is n_1 at q_1.

f is injective: Let $\rho = (q_1, n_1) \to \ldots \to (q_m, n_m), \rho' = (q'_1, n'_1) \to \ldots \to (q'_m, n'_m)$ be two paths in $\mathcal{A}$ such that $\rho \neq \rho'$. If $|\rho| \neq |\rho'|, f(\rho) \neq f(\rho')$ clearly. Suppose $|\rho| = |\rho'|$ and consider the smallest i such that either $n_i \neq n'_i$ or $q_i \neq q'_i$. If $q_i \neq q'_i$, then clearly $f(\rho) = (q_1 \to \ldots \to q_i \to \ldots \to q_m, n_1) \neq (q_1 \to \ldots \to q'_i \to \ldots \to q_m, n'_1) = f(\rho')$. Otherwise, if $q_i = q'_i$ and $n_i \neq n'_i$, note that $i = 1$: there can only be one transition from $q_{i-1} \to q_i$ in $\mathcal{A}^*$ and because i is the smallest such i, $q_{i-1} = q'_{i-1}$. Thus if $i > 1$, this would mean that $n_i = n'_i$, which is impossible. So $f(\rho) = (q_1 \to \ldots \to q_m, n_1) \neq (q'_1 \to \ldots \to q'_m, n'_1) = f(\rho')$.

f is surjective: Let $(\rho^*, n_1) = (q_1 \to \ldots \to q_m, n_1) \in \Psi^* \times [N]$. Let n_i be the value of n in $\mathcal{A}^*$ after starting at state q_1 with $n = n_1$ and traversing each state q_i in order. Then $\rho = (q_1, n_1) \to \ldots (q_m, n_m)$ is a path in $\mathcal{A}$ by construction and clearly $f(\rho) = (\rho^*, n_1)$.

Fix $x \in \mathbb{R}$ and $\rho \in \Psi$. Let $f(\rho) = (\rho^*, n_1)$. We will show that $\mathbb{P}[\epsilon, x, \rho] = \mathbb{P}[\epsilon, N, x, n_1, \rho^*]$.

This follows by induction on $|\rho^*|$:

Suppose $|\rho^*| = 0$. Then $\mathbb{P}[\epsilon, x_0, \rho] = \mathbb{P}[\epsilon, N, x_0, n_0, \rho^*] = 1$.

Now suppose $|\rho^*| = k > 0$ and that for all $|\rho'^*| < k$, $\mathbb{P}[\epsilon, x_0, \rho'] = \mathbb{P}[\epsilon, N, x_0, n_0, \rho'^*]$.

Let c_0 be the guard of the first transition $q_0 \to q_1$ in $\rho^* = q_0 q_1 \cdots q_{m-1}$. So $\delta^*(q_0, c_0) = (q_1, \sigma, b, i)$.

Let $\nu = \mu + a_0$, where a_0 is the first input value read. If $o_0 = (y, v, w)$ for $y \in \{\texttt{insample}, \texttt{insample'}\}$, then let

$$k = \int_v^w \frac{d\epsilon}{2} e^{-d\epsilon|z-\mu-a_0|} dz$$
$$k' = \int_v^w \frac{d'\epsilon}{2} e^{-d'\epsilon|z-\mu'-a_0|} dz$$

Case 1: $c = n < N$

[2]Does this need to be elaborated on?

Note that $n_0 < N$ since ρ^* is a valid path.

By construction, $\delta((q_0, n_0), \texttt{true}) = ((q_1, n_1), \sigma, b)$ where $n_1 = n_0 + i$. Let x' be the value of x at q_1 in $\mathcal{A}^*$. Since $\mathcal{A}^*$ assigns to x iff $\mathcal{A}$ does, x' is also the value of x at (q_1, n_1) in $\mathcal{A}$.

Since $n_0 < N$, by the induction hypothesis

$$\mathbb{P}[\epsilon, N, x, n_0, \rho^*] = \mathbb{P}[\epsilon, N, x', n_1, \texttt{tail}(\rho^*)] = \mathbb{P}[\epsilon, x', \texttt{tail}(\rho)] = \mathbb{P}[\epsilon, x, \rho]$$

Case 2: $c = \texttt{true}$

As in case 1, by construction, $\delta((q_0, n_0), \texttt{true}) = ((q_1, n_1), \sigma, b)$ where $n_1 = n_0 + i$. Let x' be the value of x at q_1 in $\mathcal{A}^*$. Since $\mathcal{A}^*$ assigns to x iff $\mathcal{A}$ does, x' is also the value of x at (q_1, n_1) in $\mathcal{A}$.

Then by the induction hypothesis

$$\mathbb{P}[\epsilon, N, x, n_0, \rho^*] = \mathbb{P}[\epsilon, N, x', n_1, \texttt{tail}(\rho^*)] = \mathbb{P}[\epsilon, x', \texttt{tail}(\rho)] = \mathbb{P}[\epsilon, x, \rho]$$

Case 3: $c = n \geq N$

Note that $n_0 \geq N$ since ρ^* is a valid path.

By construction, $\delta((q_0, n_0), \texttt{true}) = ((q_1, n_1), \sigma, b)$ where $n_1 = n_0$. Let x' be the value of x at q_1 in $\mathcal{A}^*$. Since $\mathcal{A}^*$ assigns to x iff $\mathcal{A}$ does, x' is also the value of x at (q_1, n_1) in $\mathcal{A}$.

Since $n_0 \geq N$, by the induction hypothesis

$$\mathbb{P}[\epsilon, N, x, n_0, \rho^*] = \mathbb{P}[\epsilon, N, x', n_1, \texttt{tail}(\rho^*)] = \mathbb{P}[\epsilon, x', \texttt{tail}(\rho)] = \mathbb{P}[\epsilon, x, \rho]$$

Case 4: $c = n < N$ **AND** $\texttt{insample} \geq \texttt{x}$

Since ρ^* is valid, $n_0 < N$.

By construction, $\delta((q_0, n_0), \texttt{insample} \geq \texttt{x}) = ((q_1, n_1), \sigma, b)$ where $n_1 = n_0 + i$.

Suppose $b = \texttt{true}$ (i.e. $\texttt{trans}(q_0)$ is an assignment transition), then:

If σ is of the form $(\texttt{insample'}, v, w)$, since $n_0 < N$,

$$\begin{aligned}
\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|} \right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)] dz \\
&= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|} \right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)] dz \text{ by the induction hypothesis} \\
&= \mathbb{P}[\epsilon, x, \rho]
\end{aligned}$$

where d, ν, k' are defined as in section 1.1.

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)]dz \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

Now suppose that $b = \texttt{false}$. If σ is of the form $(\texttt{insample'}, v, w)$, since $n_0 < N$,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_{x}^{\infty} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)] \\ &= k' \left(\int_{x}^{\infty} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k' are defined as in section 1.1[3].

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k', ℓ are defined as in section 1.1.

Case 5: $c = n < N$ **AND** $\texttt{insample} < \texttt{x}$

Since ρ^* is valid, $n_0 < N$.

By construction, $\delta((q_0, n_0), \texttt{insample} < \texttt{x}) = ((q_1, n_1), \sigma, b)$ where $n_1 = n_0 + i$.

Suppose $b = \texttt{true}$. Then if σ is of the form $(\texttt{insample'}, v, w)$, since $n_0 < N$,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_{-\infty}^{x} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)]dz \\ &= k' \left(\int_{-\infty}^{x} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)]dz \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k' are defined as in section 1.1.

[3]todo: fix these

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)]dz \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

Now suppose that $b = \texttt{false}$. If σ is of the form $(\texttt{insample'}, v, w)$, since $n_0 < N$,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|} dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)] \\ &= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|} dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k' are defined as in section 1.1[4].

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|} dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|} dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k', ℓ are defined as in section 1.1.

Case 6: $c = \texttt{insample} \geq \texttt{x}$

By construction, $\delta((q_0, n_0), \texttt{insample} \geq \texttt{x}) = ((q_1, n_1), \sigma, b)$ where $n_1 = n_0 + i$.

Suppose $b = \texttt{true}$ (i.e. $\texttt{trans}(q_0)$ is an assignment transition), then:

If σ is of the form $(\texttt{insample'}, v, w)$,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)]dz \\ &= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)]dz \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k' are defined as in section 1.1.

[4]todo: fix these

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)]dz \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

Now suppose that $b = \texttt{false}$. If σ is of the form $(\texttt{insample'}, v, w)$,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_{x}^{\infty} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)] \\ &= k' \left(\int_{x}^{\infty} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k' are defined as in section 1.1[5].

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_{\max(x,\ell)}^{u} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k', ℓ are defined as in section 1.1.

Case 7: $c = \texttt{insample} < \texttt{x}$

By construction, $\delta((q_0, n_0), \texttt{insample} < \texttt{x}) = ((q_1, n_1), \sigma, b)$ where $n_1 = n_0 + i$.

Suppose $b = \texttt{true}$. Then if σ is of the form $(\texttt{insample'}, v, w)$,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_{-\infty}^{x} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)]dz \\ &= k' \left(\int_{-\infty}^{x} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)]dz \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k' are defined as in section 1.1.

[5]todo: fix these

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, N, z, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}\right) \mathbb{P}[\epsilon, z, \texttt{tail}(\rho)]dz \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

Now suppose that $b = \texttt{false}$. If σ is of the form $(\texttt{insample'}, v, w)$,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)] \\ &= k' \left(\int_x^\infty \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k' are defined as in section 1.1[6].

Otherwise,

$$\begin{aligned}\mathbb{P}[\epsilon, N, x, n_0, \rho^*] &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, N, x, n_1, \texttt{tail}(\rho^*)]dz \\ &= \left(\int_\ell^{\min(u,x)} \frac{d\epsilon}{2} e^{-d\epsilon|z-\nu|}dz\right) \mathbb{P}[\epsilon, x, \texttt{tail}(\rho)] \text{ by the induction hypothesis} \\ &= \mathbb{P}[\epsilon, x, \rho]\end{aligned}$$

where d, ν, k', ℓ are defined as in section 1.1.

This is sufficient to prove the lemma.

□

Lemma 2.6. *There exists $d > 0$ such that $\mathcal{A}$ is $d\epsilon-$differentially private if and only if $\mathcal{A}^*$ is $d\epsilon$-differentially private.*

Proof. Let f be a bijection from paths in $\mathcal{A}$ to tuples of paths in $\mathcal{A}^*$ to $[N]$, as defined in Lemma 2.5.

Suppose that $\exists d > 0$ such that $\mathcal{A}$ is $d\epsilon-$differentially private. Then for all equivalent paths ρ, ρ' in $\mathcal{A}$ from the start state $(q_0, 0)$ such that $\texttt{inseq}(\rho)$ and $\texttt{inseq}(\rho')$ are adjacent, $\mathbb{P}[\epsilon, \rho] \leq e^{d\epsilon}\mathbb{P}[\epsilon, \rho']$. Consider two equivalent paths ρ^*, ρ'^* in $\mathcal{A}^*$ from the start state q_0 such that $\texttt{inseq}(\rho^*)$ and $\texttt{inseq}(\rho'^*)$ are adjacent. Then $\forall x \in \mathbb{R}, \mathbb{P}[\epsilon, N, x, 0, \rho^*] = \mathbb{P}[\epsilon, x, f^{-1}((\rho^*, 0))] \leq e^{d\epsilon}\mathbb{P}[\epsilon, x, f^{-1}((\rho'^*, 0))] = e^{d\epsilon}\mathbb{P}[\epsilon, N, x, 0, \rho'^*]$. Thus, $\mathcal{A}^*$ is $d\epsilon$-differentially private.

[6]todo: fix these

Suppose that $\forall d > 0, \mathcal{A}$ is not $d\epsilon$-differentially private. So there exists two equivalent paths from the start state $(q_0, 0)$ in $\mathcal{A}$ $\rho = (q_0, 0) \to \ldots \to (q_m, n_m), \rho' = (q_0, 0) \to (q'_m, n'_m)$ in $\mathcal{A}$ such that $\texttt{inseq}(\rho)$ and $\texttt{inseq}(\rho')$ are adjacent, but $\mathbb{P}[\epsilon, \rho] > e^{d\epsilon}\mathbb{P}[\epsilon, \rho']$.

Let $f(\rho) = (\rho^*, 0)$ and $f(\rho') = (\rho'^*, 0)$. Fix $\texttt{inseq}(\rho^*) = \texttt{inseq}(\rho)$ and $\texttt{inseq}(\rho'^*) = \texttt{inseq}(\rho')$.

Then by Lemma 2.5, $\mathbb{P}[\epsilon, N, \rho] = \mathbb{P}[\epsilon, N, x, 0, \rho^*] = \mathbb{P}[\epsilon, x, \rho] = \mathbb{P}[\epsilon, \rho] > e^{d\epsilon}\mathbb{P}[\epsilon, \rho'] = e^{d\epsilon}\mathbb{P}[\epsilon, x, \rho'] = e^{d\epsilon}\mathbb{P}[\epsilon, N, x, 0, \rho'^*] = e^{d\epsilon}\mathbb{P}[\epsilon, N, \rho'^*]$. Thus, $\mathcal{A}^*$ is not $d\epsilon$-DP. □

Lemmas 2.3 and 2.6 together prove the theorem. □

Corollary 2.7. *Let $\mathcal{A}^*$ be a DiPA* with unfixed parameters.*

Let $f(\epsilon, N) : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ be defined as follows:

Consider the instantiated version of $\mathcal{A}^*$ with parameters ϵ and N. Let $\mathcal{A}$ be the DiPA constructed from $\mathcal{A}^*$ as in Theorem 3.1. $f(\epsilon, N) = wt(\mathcal{A})$.

Then $\forall \epsilon, f(\epsilon, N)$ grows linearly in N.

Corollary 2.8. *For a DiPA* $\mathcal{A}^*$, the well-formedness of $\mathcal{A}^*$ can be decided efficiently.*

Proof. Note that the time it takes is bounded by the cost of creating a DiPA $\mathcal{A}$ from $\mathcal{A}^*$ as in Theorem 3.1. The construction of $\mathcal{A}$ from $\mathcal{A}^*$ causes the number of states to increase by a factor of N. Each transition in $\mathcal{A}^*$ corresponds to at most N transitions in $\mathcal{A}$. Since the well-formedness of $\mathcal{A}$ and $\mathcal{A}^*$ are equivalent, at most there is a linear increase in the time required to check the well-formedness of $\mathcal{A}^*$ as compared to a DiPA* of the same size. □

References

[1] Rohit Chadha, A. Prasad Sistla, and Mahesh Viswanathan. On Linear Time Decidability of Differential Privacy for Programs with Unbounded Inputs, April 2021.