

# 1 DiPA\*

**Definition 1.1.** Fix parameters  $\epsilon, N$ . Let  $C$  be the guard conditions  $\{n < N, \text{true}, \text{insample} \geq x, \text{insample} < x, n < N \text{ AND } \text{insample} \geq x, n < N \text{ AND } \text{insample} < x\}$ . 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  $Q_{non}$
- $\Sigma$  is the input alphabet (taken to be  $\mathbb{R}$ )
- $\Gamma$  is a finite output alphabet
- $q_0 \in Q$  is the starting state
- $X = \{x, \text{insample}, \text{insample}', n\}$  is a set of variables.  $x, \text{insample}, \text{insample}' \in \mathbb{R}$ ;  $n \in \mathbb{N}$  and is initialized to 0.
- $P : Q \rightarrow \mathbb{Q}^{\geq 0} \times \mathbb{Q} \times \mathbb{Q}^{\geq 0} \times \mathbb{Q}$  describing the parameters for sampling from Laplace distributions at each state.
- $\delta : (Q \times C) \rightarrow Q \times (\Gamma \cup \{\text{insample}, \text{insample}'\} \cup \{\phi\}) \times \{\text{true}, \text{false}\} \times \{0, 1\}$  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  $\delta$  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  $\delta(q, \text{true})$  is defined, then  $\delta(q, n < N)$  cannot be defined and vice versa. Similarly, at most one of  $\delta(q, \text{insample} \geq x)$  and  $\delta(q, n < N \text{ AND } \text{insample} \geq x)$  can be defined and at most one  $\delta(q, \text{insample} < x)$  and  $\delta(q, n < N \text{ AND } \text{insample} < x)$  can be defined. For any state  $q \in Q$ , if  $\delta(q, \text{true})$  is defined, then  $\delta(q, \text{insample} \geq x)$  and  $\delta(q, \text{insample} < x)$  are not defined.

Similarly, if  $\delta(q, n < N)$  is defined, then  $\delta(q, n < N \text{ AND } \text{insample} \geq x)$  and  $\delta(q, n < N \text{ AND } \text{insample} < x)$  are not defined. Finally, if  $\delta(q, n \geq N)$  is defined, then  $\delta(q, \text{true})$ ,  $\delta(q, \text{insample} \geq x)$  and  $\delta(q, \text{insample} < x)$  are not defined.

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

- **Output Distinction:** For any state  $q \in Q$ , if  $\delta(q, \text{insample} \geq x) = (q_1, o_1, b_1, i_1)$  and  $\delta(q, \text{insample} < x) = (q_2, o_2, b_2, i_2)$ , 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  $\phi$  output symbol is reserved for transitions with guard  $n \geq N$ , which must output  $\phi$ .
- **Initialization:** The initial state  $q_0$  has only one outgoing transition of the form  $\delta(q_0, \text{true}) = (q, o, \text{true}, i)$  for  $i \in \{0, 1\}$ .

- **Non-input transition:** From any  $q \in Q_{non}$ , if  $\delta(q, c)$  is defined, then  $c = \mathbf{true}$ .
- **Control Flow Separation:** Consider the underlying graph  $G$  of  $\mathcal{A}$ . For all states  $q \in Q$ , if  $\delta(q, n \geq N) = (q', o, b, i)$ , let  $G'$  be the graph of  $G$  with the edge corresponding to the transition with guard  $n \geq N$  from  $q$  removed. Then  $q$  and  $q'$  must be in different 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 Path Probabilities

**Definition 1.2.** (summarized from [1]) A **path**  $\rho$  of length  $n$  of a DiPA\*  $\mathcal{A}$  is a sequence of states, inputs, and outputs  $\rho = q_0 \xrightarrow{a_0, o_0} q_1 \rightarrow \dots \rightarrow q_{n-1}$ , 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  $q_i \rightarrow q_{i+1}$ . We denote the sequence of inputs  $a_i$  for a path  $\rho$  as  $\mathbf{inseq}(\rho)$  and the sequence of outputs  $o_i$  as  $\mathbf{outseq}(\rho)$ . In general, for a path  $\rho = q_0 \rightarrow q_1 \rightarrow \dots \rightarrow q_{n-1}$  we denote the transition  $q_i \rightarrow q_{i+1}$  by  $\mathbf{trans}(q[i])$  and the guard of  $\mathbf{trans}(\rho[i])$  as  $\mathbf{guard}(\rho[i])$ .

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

For any path  $\rho$  of a DiPA\*  $\mathcal{A}$ , we define  $\mathbb{P}[\epsilon, N, x, n, \rho]$  as the **probability** of path  $\rho$  being traversed with  $\mathcal{A}$  parameters  $\epsilon$  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  $\rho = q_0 \xrightarrow{a_0, o_0} q_1 \rightarrow \dots \xrightarrow{a_{n-1}, o_{n-1}} q_n$ . Here,  $a_i$  and  $o_i$  are the input to state  $q_i$  and output of transition  $q_i \rightarrow q_{i+1}$ , 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 **insample** and **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  $y \in \{\mathbf{insample}, \mathbf{insample}'\}$ , 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  $\rho$  is not an assignment transition (i.e.  $b = \mathbf{false}$ ), 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  $\mathbb{P}[\epsilon, N, x, n, \rho] = \mathbb{P}[\epsilon, N, x, n + i, \text{tail}(\rho)]$ . If  $o_0 = (\text{insample}, v, w)$  then  $\mathbb{P}[\epsilon, N, x, n, \rho] = k\mathbb{P}[\epsilon, N, x + i, \text{tail}(\rho)]$ . If  $o_0 = (\text{insample}, v, w)$  then  $\mathbb{P}[\epsilon, N, x, n, \rho] = k'\mathbb{P}[\epsilon, N, x, n + i, \text{tail}(\rho)]$

**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  $\mathbb{P}[\epsilon, N, x, n, \text{tail}(\rho)]$  is referenced in [1],  $\mathbb{P}[\epsilon, N, x, n + i, \text{tail}(\rho)]$  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  $\rho$  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  $\rho = q_0 \rightarrow q_1 \rightarrow \dots q_n$  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  $\epsilon, N$  is  $d\epsilon$ -**differentially private** if for all equivalent paths  $\rho, \rho'$  in  $\mathcal{A}$  such that  $\text{inseq}(\rho)$  and  $\text{inseq}(\rho')$  are adjacent,  $\mathbb{P}[\epsilon, N, \rho] \leq e^{d\epsilon}\mathbb{P}[\epsilon, N, \rho']$ .

## 2 Violations of Differential Privacy

**Definition 2.1.** An **abstract path** in a DiPA\* is a sequence of states and associated outputs  $q_0\sigma_0q_1\sigma_1 \dots q_{n-1}\sigma_{n-1}q_n$ .

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

**Definition 2.2.** 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 2.3.** A cycle  $C$  with an exit state with transition  $n \geq N$  is an **infeasible** cycle if, for *all* paths  $\rho = q_0 \rightarrow q_1 \rightarrow \dots \rightarrow q_m$  from the start state to a state  $q_m \in C$ , at least  $N$  transitions  $q_i \rightarrow q_{i+1}$  are increment transitions or some transition  $q_i \rightarrow q_{i+1}$  has guard  $n \geq N$ . Otherwise,  $C$  is **feasible**.

**Definition 2.4.** (from [1]) A **leaking cycle** is a cycle  $C = q_0 \xrightarrow{a_0, o_0} q_1 \rightarrow \dots \rightarrow q_{n-1} \rightarrow q_0$  in a DiPA\*  $\mathcal{A}$  if there exist indices  $0 \leq i < j < n$  such that the  $i$ th transition  $q_i \rightarrow q_{i+1}$  is an assignment transition and the guard of the  $j$ th transition guard is not  $n < N$ .

**Proposition 2.5.** *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  $\eta_\ell$  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  $\eta$  with the cycle  $C$  repeated  $\ell$  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  $\eta_\ell$  in  $\mathcal{A}$  exists for all  $\ell > 0$ . Thus, the same input sequences  $\alpha_\ell$  and  $\beta_\ell$  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  $\eta_\ell$  has at most  $f$  transitions with guard  $n \geq N$ , even as  $\ell$  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  $\eta_\ell$  in  $\mathcal{A}$  exists for all  $\ell > 0$ , so  $\alpha_\ell$  and  $\beta_\ell$  from Lemma 6 of [1] are witnesses to a violation of differential privacy.

□

**Definition 2.6.** (from [1]) A cycle  $\rho$  of a DiPA\*  $\mathcal{A}$  is an **L-cycle** (respectively, **G-cycle**) if there is an  $i < |\rho|$  such that  $\text{guard}(\rho[i]) = \text{insample} < \mathbf{x}$  (respectively  $\text{guard}(\rho[i]) = \text{insample} \geq \mathbf{x}$ ).

**Definition 2.7.** (from [1]) A path  $\rho$  of a DiPA  $\mathcal{A}^*$  is an **AL-path** (respectively, **AG-path**) if all assignment transitions on  $\rho$  have guard  $\text{insample} < \mathbf{x}$  (respectively,  $\text{insample} \geq \mathbf{x}$ ).

**Definition 2.8.** (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 2.9.** 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  $\rho$  or  $C'$  has no exit state.
2. Either there exists some path  $\tau$  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  $\tau$  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  $\rho$  from states  $a_i$  to  $a_{i+1}$ , there exists some path  $\tau$  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  $\tau$ .

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  $\rho$  between  $C$  and  $C'$  is in fact traversible.

**Proposition 2.10.** *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

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

for all  $\ell > 0$ . Let  $t_k$  be the  $k$ -th transition of  $\eta$  and  $c_k$  be the guard of  $t_k$ .  $\nu_\ell$  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  $c_j = \text{insample} \geq x$

Intuitively,  $\forall \ell > 0$ ,  $\eta_\ell$  is a path that begins at the start state, reaches  $C$ , traverses  $C$   $\ell$  times, then traverses an **AG**-path to  $C'$  before traversing  $C'$   $\ell$  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  $\gamma$  and  $\gamma'$  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  $\rho$  that have guard  $n \geq N$ , then  $s \cdot \gamma^\ell \cdot \rho$  is valid for all  $\ell > 0$ .

If there exists a transition  $\rho$ , then by condition (4) from the definition of feasible unbounded leaking pairs, for all such transitions  $a_i \rightarrow a_{i+1}$ , 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  $\rho'$  is a subpath of  $\rho$  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**  $\geq \mathbf{x}$ , or **insample**  $< \mathbf{x}$ , which means that a path can traverse  $C'$   $\ell$  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  $\alpha$  is a path from the start state to  $a_1$  such that there are no more than  $N - 1$  increment transitions in  $\tau$ . Notably, this path is valid because  $C$  is reachable and because  $\rho$  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 **true**, **insample**  $\geq \mathbf{x}$ , or **insample**  $< \mathbf{x}$ . In addition, because when following such a path  $\tau$ ,  $n < N$  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  $\ell > 0$ .  $\square$

**Definition 2.11.** (from [1]) A cycle  $C$  of a DiPA  $\mathcal{A}$  is a **disclosing cycle** if there exists some  $0 \leq i < |C|$  such that  $\text{trans}(C[i])$  is an input transition that outputs either **insample** or **insample'**.

**Proposition 2.12.** *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  $\eta_\ell$  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  $\eta$  with the cycle  $C$  repeated  $\ell$  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  $\eta_\ell$  in  $\mathcal{A}$  exists for all  $\ell > 0$ . Thus, the same input sequences  $\alpha_\ell$  and  $\beta_\ell$  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  $\eta_\ell$  has at most  $f$  transitions with guard  $n \geq N$ , even as  $\ell$  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  $\eta_\ell$  in  $\mathcal{A}$  exists for all  $\ell > 0$ , so  $\alpha_\ell$  and  $\beta_\ell$  from Lemma 8 of [1] are witnesses to a violation of differential privacy.  $\square$

**Definition 2.13.** (adapted from [1]) An feasible unbounded privacy violating lasso is a path  $\rho = a_1 \cdots a_k$  of length  $n$  in a DiPA\*  $\mathcal{A}$  such that one of the following hold:

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

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

In addition, if there are any transitions  $a_i \rightarrow a_{i+1}$  in  $\rho$  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  $\alpha$  represents a path from the start state of  $\mathcal{A}$  to  $a_1$  and  $\beta$  represents a subpath of  $\rho$  from  $a_1$  to  $a_i$ .

**Definition 2.14.** For a lasso  $\rho$ , let  $C_\rho$  be the cycle associated with<sup>1</sup>  $\rho$ . Then a lasso  $\rho$  in a DiPA\*  $\mathcal{A}$  is bounded iff  $C_\rho$  is bounded. Similarly,  $\rho$  is feasible iff  $C_\rho$  is feasible.

**Proposition 2.15.** *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  $\text{first}(\rho) = a_1$  is in an L-cycle  $C$  where the last transition of  $\rho$  has guard  $\text{insample} \geq \mathbf{x}$  and outputs  $\text{insample}$ .

Similar to propositions 2.4, 2.10, and 2.12, we will construct a valid path  $\eta_\ell$  in  $\mathcal{A}$  starting at the start state that traverses  $C$   $\ell$  times and then traverses the path  $\rho$  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  $\gamma$  be the word representing  $C$ .

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

**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  $\ell > 0$ .

**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 2.16.** 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 2.17.** *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.

□

---

<sup>1</sup>Hopefully this is clear



### 3 Proving Differential Privacy

**Theorem 3.1.** *If a DiPA\* is well-formed, then it is differentially private.*

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

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

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, then for all  $k \in [N - 1]$ , define the transition

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

If  $\delta^*(q, n \geq N) = (q', \sigma, \mathbf{b}, x)$  is defined, then define the transition

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

Otherwise if  $\delta^*(q, g) = (q', \sigma, \mathbf{b}, x)$  is defined, then define the transition

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

All other transitions are undefined. Note that we need to separate out the  $k = N$  case to satisfy the condition of determinism.

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  $P((q, k)) = P^*(q)$ .

**Lemma 3.2.** *If  $\mathcal{A}^*$  is well-formed, then  $\mathcal{A}$  is well-formed.*

*Proof.* We will prove this using the following lemma:

**Lemma 3.3.** *If there exists a reachable cycle  $C = (a_0, k_0) \rightarrow (a_1, k_1) \rightarrow \dots \rightarrow (a_{m-1}, k_{m-1}) \rightarrow (a_0, k_0)$  in  $\mathcal{A}$ , then there exists a reachable unbounded feasible cycle  $C^* = a_0 \rightarrow a_1 \rightarrow \dots \rightarrow a_{m-1} \rightarrow a_0$  in  $\mathcal{A}^*$ .*

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

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

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  $(a_i, k_i) \rightarrow$

$(a_{i+1}, k_i + 1)$ , 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  $k_0 = k_1 = \dots k_{m-1} = n$  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  $(s, n) \rightarrow (s', n)$  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.  $\square$

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

**Lemma 3.4.** *If  $\mathcal{A}$  is  $wt(\mathcal{A})$ -DP, then  $\mathcal{A}^*$  is  $wt(\mathcal{A})$ -DP.*

*Proof.* Let  $\rho^* = q_0 q_1, \dots, q_{m-1}$  be a path of length  $m$  in  $\mathcal{A}^*$ . Let  $n_i$  be the value of the variable  $n$  in  $\mathcal{A}^*$  at state  $q_i$  (so e.g.  $n_0 = 0$ ). Then  $\rho = (q_0, n_0)(q_1, n_1) \dots (q_{m-1}, n_{m-1})$  is a path in  $\mathcal{A}$ . This follows immediately from the construction of  $\mathcal{A}$ . In addition,  $\mathbb{P}[\epsilon, \rho] = \mathbb{P}[\epsilon, N, \rho^*]$ .

This also follows immediately by induction on  $|\rho|$ .

This is sufficient to prove the lemma.  $\square$

Lemmas 3.2 and 3.4 together prove the theorem.  $\square$

**Corollary 3.5.** *Let  $\mathcal{A}^*$  be a DiPA<sup>\*</sup> with unfixed parameters.*

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

*Consider the instantiated version of  $\mathcal{A}^*$  with parameters  $\epsilon$  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 3.6.** *For a DiPA<sup>\*</sup>  $\mathcal{A}^*$ , 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.  $\square$

## References

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