

Differentially Private Event Logs for Process Mining: Supplementary Material

Gamal Elkoumy¹, Alisa Pankova², and Marlon Dumas¹

¹ University of Tartu, Tartu, Estonia
{gamal.elkoumy,marlon.dumas}@ut.ee

² Cybernetica, Tartu, Estonia
{alisa.pankova}@cyber.ee

Abstract. In this document, we provide supplementary material for our paper. We list the selected event logs and their characteristics and descriptive statistics. Also, we list the event logs attached with the document as the result of the experiments.

Keywords: Process Mining · Privacy-Enhancing Technologies · Differential Privacy

1 Evaluation Supplementary Material

1.1 Dataset Selection

Table 1 shows the selected event logs and their descriptive statistics.

Table 1: Descriptive Statistics of Event Logs

event log	# Traces	# Tasks	# Events	# Edges	Case Variant	Trace Length		Case Duration		
						Min	Max	Min	Max	Avg
<i>BPI12</i> [5]	13087	23	262200	116	4366	3	175	1.85 s	4.51 m	1.23 w
<i>BPI13_i</i> [14]	7554	4	65533	16	1511	1	123	inst.	2.11 y	1.73 w
<i>BPI14_i</i> [2]	46616	39	466737	497	22632	1	178	14 s	1.07 y	5.07 d
<i>BPI15₁</i> [3]	1199	398	52217	495	1170	2	101	8.56 h	4.07 y	3.15 m
<i>BPI17</i> [6]	31509	24	1202267	181	3942	10	180	3.35 m	9.4 m	3.13 w
<i>BPI18</i> [8]	43809	14	2514266	499	28457	24	2973	3.74 m	2.77 y	11.03 m
<i>BPIC19</i> [7]	251734	42	1595923	498	11973	1	990	2 ms	70.33 y	2.35 m
<i>BPI20_r</i> [4]	7065	51	86581	500	1478	3	90	12.61 h	3.26 y	2.87 m
<i>CCC19</i> [13]	20	29	1394	149	20	52	118	11 m	1.01 d	1.73 h
<i>CredReq</i> [1]	10035	8	150525	9	1	15	15	3.5 h	5 d	22 h
<i>Hospital</i> [12]	1143	624	150291	903	981	1	1814	inst.	3.17 y	1.06 y
<i>Sepsis</i> [11]	1050	16	15214	115	846	3	185	2.03 m	1 y	4 w
<i>Traffic</i> [10]	150370	11	561470	77	231	2	20	3 d	12 y	11 m
<i>Unrine.</i> [9]	1650	10	6973	25	50	2	35	10.1 m	2.32 y	3.7 w

1.2 Folder Organization

The attached folder to this PDF contains the anonymized event logs and the error metrics. The error metrics are presented as a CSV file, called “combined_error_metrics.csv”. The anonymized event logs are organized into folders. The high level folders is the δ values for the anonymized event logs. The sub-folders represent the value of the precision parameter p . The naming convention of the files are of the form “e_anonymized_s1_t.s2.s3.s4”. “s1” represents the event log name. “s2” represents the value of the precision parameter. “s3” represents the value of the δ parameter. “s4” represents the iteration. The ϵ values used for every event and for every case variants is stored in the event log. The event attribute “epsilon_per_event” represents the ϵ used to anonymize the time component of this event. The case attribute “epsilon_per_trace” represents the ϵ value used to anonymize the frequency of all the instances of this case variant.

References

1. Djedović, A.: Credit Requirement Event Logs (9 2017). <https://doi.org/10.4121/uuid:453e8ad1-4df0-4511-a916-93f46a37a1b5>
2. van Dongen, B.B.: Bpi challenge 2014 . 4tu.researchdata. collection. (2014). <https://doi.org/10.4121/uuid:c3e5d162-0cfd-4bb0-bd82-af5268819c35>
3. van Dongen, B.B.: Bpi challenge 2015 . 4tu.researchdata. collection. (2015). <https://doi.org/10.4121/uuid:31a308ef-c844-48da-948c-305d167a0ec1>
4. van Dongen, B.B.: Bpi challenge 2020 . 4tu.researchdata. collection. (2020). <https://doi.org/10.4121/uuid:52fb97d4-4588-43c9-9d04-3604d4613b51>
5. van Dongen, B.: BPI Challenge 2012 (4 2012). <https://doi.org/10.4121/uuid:3926db30-f712-4394-aebc-75976070e91f>
6. van Dongen, B.: BPI Challenge 2017 (2 2017). <https://doi.org/10.4121/uuid:5f3067df-f10b-45da-b98b-86ae4c7a310b>
7. van Dongen, B.: BPI Challenge 2019 (1 2019). <https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1>
8. van Dongen, B., Borchert, F.F.: BPI Challenge 2018 (3 2018). <https://doi.org/10.4121/uuid:3301445f-95e8-4ff0-98a4-901ff1f204972>, https://data.4tu.nl/articles/dataset/BPI_Challenge_2018/12688355
9. Gunst, P.: Urineweginfectie (UWI-casus) logboek (8 2020). <https://doi.org/10.4121/12826727.v1>
10. de Leoni, M.M., Mannhardt, F.: Road Traffic Fine Management Process (2 2015). <https://doi.org/10.4121/uuid:270fd440-1057-4fb9-89a9-b699b47990f5>
11. Mannhardt, F.: Sepsis cases-event log. Eindhoven University of Technology. Dataset pp. 227–228 (2016)
12. Mannhardt, F.: Hospital Billing - Event Log (8 2017). <https://doi.org/10.4121/uuid:76c46b83-c930-4798-a1c9-4be94dfef741>
13. Munoz-Gama, J., de la Fuente, R.R., Sepúlveda, M.M., Fuentes, R.R.: Conformance Checking Challenge 2019 (CCC19) (2 2019). <https://doi.org/10.4121/uuid:c923af09-ce93-44c3-ace0-c5508cf103ad>
14. Steemanvan, W.: Bpi challenge 2013 . 4tu.researchdata. collection. (2014). <https://doi.org/10.4121/uuid:a7ce5c55-03a7-4583-b855-98b86e1a2b07>