1 Proof

Lemma 1.1 (The modified Heuristic Miner is $n\epsilon_1 \cdot \frac{cap}{\tau \# traces}$ Differential Private)**.**

Proof. We apply n times the AboveThreshold-Mechanism with noise parameter ϵ_1, to all elements in the Event-Log, so that we can exclude edges that were visited in less than τ number of traces in the Event-log D.

For the ith invocation of AboveThreshold, we have a sequence $[(\perp)_{i=1,\ldots,m_i-1}, \top]$ (for some $m_i \in \mathbb{N}$). We then compute BelowThreshold for the same queries and noise parameter ϵ_1, until we get a $\perp$, so a sequence looks like $[(\top)_{i=1,\ldots,m_{i+1}-1}, \perp]$. So if we get $\top$ from the AboveThreshold-Mechanism, we have to get $\perp$ for that same query with the BelowThreshold-Mechanism. The BelowThreshold-Mechanism provides the same privacy as the AboveThreshold-Mechanism, which can be proven like the AboveThreshold-Mechanism in [1]. These results can be used without any additional leakage, due to the post-processing theorem.

From [1] we know that the AboveThreshold-Mechanism is ϵ-DP, with ϵ being the noise parameter. With the use of the basic composition theorem we also know that executing a ϵ-DP algorithm n_1 times, we get that it is $n_1\epsilon$-DP. Therefore executing the AboveThreshold-Mechanism n times, with noise parameter ϵ_1, we get that $n\epsilon_1$-DP holds.

Next, the algorithm compute the frequency for every edge that evaluated with $\top$. If an edge has a frequency greater than *cap*, then we will cap it at *cap*. Hence, we minimize leakage about which processes are carried out the most. Additionally, we filter out edges that occur in less than a $\tau \in [0, 1]$ fraction of traces in the Event-Log. We consider the queries of the following form: $q_e(D) :=$ "In which frequency does edge e to occur in the Event-log D?" Each of these queries has sensitivity $cap/(\tau \# traces)$.

Therefore $n\epsilon_1 \cdot \frac{cap}{\tau \# traces}$-DP holds, due to the basic composition theorem. Everything that is done with this graph has no additional leakage due to the post-processing theorem. □

References

[1] Cynthia Dwork, A.R.: The Algorithmic Foundations of Differential Privacy (2014), `https://www.cis.upenn.edu/ aaroth/Papers/privacybook.pdf`