Federated Process Mining: Extending Traditional Process Mining to Meet Confidentiality Requirements

Christian Rennert

Chair of Process and Data Science, RWTH Aachen University, Aachen, Germany

Abstract

Inter-organizational process mining creates value by giving insights in processes that span across multiple organizations. Federated process mining is a discipline within inter-organizational process mining that gives guarantees towards confidentiality, meeting the requirements of organizational secrecy and privacy regulations. However, federated variants of traditional process mining algorithms, that do not pose limitations to the control flow, are in majority missing. Further, a categorization of different confidentiality requirements between domains is missing. This work relates the existing work in federated process mining to future projects and highlights challenges in federated process mining that need to be tackled. In detail, we identify (1) a classification of potential secrets, (2) the inter-organizational variant analysis, (3) the development of new privacy-preserving inter-organizational process mining algorithms, (4) mitigation techniques from having an undesired secret sharing with the revealed results, (5) and the interaction with the mitigation techniques and partial results by users as highly relevant research challenges.

Keywords

Federated process mining, Inter-organizational, Privacy, Confidentiality, Multi-party computation

1. Introduction

Modern process mining is adapted by commercial vendors and aims to create value from an organization's event data stored in information systems. However, support for process mining on processes that include multiple organizations is missing due to the gap in research for methods that are efficient and privacy-preserving. Even thought that there are methods on inter-organizational process mining [1], they may be infeasible due to international laws on data privacy, e.g., HIPAA [2] for medical data in the US or the GDPR [3] in the EU. Furthermore, using such methods may also not be in favor of the organizations, as sharing their data for analysis might result in their data not remaining confidential.

In comparison to research on privacy and confidentiality in single organizations [4, 5], only recently has federated process mining [6] been proposed to consider confidentiality in the application of inter-organizational process mining. In [6], an abstraction-based framework for federated process mining is proposed. Abstractions are used as means to share knowledge on confidential event data without the necessity of sharing the event data itself. The abstractions can then be merged to either obtain the original result as if all confidential event data was shared, or a result that is similar to the expected result.

2. Related Work

The abstraction-based federated process mining approach is considered in later work on process discovery [7, 8, 9, 10] and conformance checking [11]. Even though abstractions enable inter-organizational process mining that meets the need for confidentiality while being performant, the currently given solutions impose restrictions on the types of inter-operability between the involved organizations: Abstraction-based process mining techniques either require (1) information on the handover-of-work to correctly reproduce the original results or (2) all activities across the partial event logs to be disjoint.

ICPM Doctoral Consortium and Demo Track 2025, October 20-24, 2025, Montevideo, Uruguay

rennert@pads.rwth-aachen.de (C. Rennert)

0000-0003-4614-6171 (C. Rennert)

Using abstractions is not suitable for all inter-organizational processes. This can be given for processes in which organizations do not know if a given case is shared or not. For example, different healthcare service providers may be interested in the hidden process that their patients follow, however, a patient may interact with many providers or only a single one, which the providers may not know or be willing to share. While being disallowed by privacy regulations, such interactions can be potentially traced by considering the patient's health insurance ID. Therefore, the requirement (1) for abstraction-based techniques cannot be fulfilled. Further, if two healthcare providers cover an overlap of their offered services, some patients may use or participate in the same or similar services at both providers, thus violating requirement (2).

Next to abstractions, confidentiality in federated process mining is addressed by using stochastic methods that approximate the traditional process mining results [12], by using fully homomorphic encryption [13] and multi-party computation [14] including a third party and allowing to interact with the results by querying [15], or by using trusted execution environments for the algorithmic computations [16].

3. Open Research Challenges

While both, abstraction-based and non-abstraction-based federated process mining approaches present promising results, a systematic approach towards federated process mining remains vague as it is not discussed yet, what information can be exposed and which could be kept as secret in our manifold process mining use cases. Therefore, we propose the following five research questions to be tackled within the next years:

RQ1. What are the different secrets that should remain confidential in federated process mining algorithms, e.g., when applying federated discovery or conformance checking?

RQ2. How to efficiently detect whether two inter-organizational traces share a case? How to identify and communicate trace variants without leaking information?

RQ3. For the results obtained from using privacy-focused algorithms: Can organizations limit critical information from being shared, by (1) being able to review the shared insights, by (2) forbidding upfront particular insights from being shared, or by (3) every organization only obtaining a share of the result that is directly related to their event data?

RQ4. How to allow domain experts, researchers, and privacy managers for interaction in cases (1) and (2) of RQ3 or when given only partial results, i.e., Case (3) of RQ3?

Given these research questions, we want to briefly highlight the relevance of each research question. RQ1 gives the basis of our work, as it defines the constraints in which research on federated process mining needs to be done. RQ2 is of importance as variant identification is crucial for the applicability of many process mining methods in terms of their scalability and performance, e.g., for process discovery, such as the discovery of process trees [17], considering only trace variants and their frequency highly improves the runtime. While being of particular interest, communicating all cases that belong to the same trace variant, may result in an attack vector as background knowledge from the different cases could be combined. However, to position federated process mining and to give a realistic chance of adoption in real-life settings or of a case study with different organizations being involved, federated process mining needs a minimal support of traditional process mining algorithms being available for federated usage. While confidentiality is important throughout the application of algorithms in federated process mining as highlighted by RQ1, RQ3 highlights that also the results of algorithms are at risk of sharing confidential information. Therefore, a need for mitigation techniques is needed that allows practitioners and researchers to guide the application. Finally, RQ4 highlights the need for new well understandable visualizations and for a framework to interact with (partial) results and parameters in a distributed manner.

4. Possible Techniques and Planned Methodology

In general, a good start to approach RQ1-RQ4 is to give a review of the existing work. A current state of inter-organizational process mining [18] and privacy and process mining [19] are given, including federated process mining. Thus, the existing reviews can pose as an overview to relate the existing work in both fields to the research questions proposed, which can be covered in future work.

Next, translating algorithms in process discovery, conformance checking and performance analysis gives examples to further investigate what information are leaked by the different algorithms and their results and what therefore can be kept secret, relating to RQ1. As a result of doing so, we can follow the examples to give first answers to RQ3 and RQ4 by giving particular examples. While investing all of that questions, we expect RQ2 to be considered as its own endeavor to be taken in parallel.

So far, our only own work contributing to our research is the contribution in [20]. In [20] we propose a directly-follows graph discovery approach that categorizes the existing discovery methods in terms of their given confidentiality and limitations on the control flow, addressing RQ1. The confidentiality of the results, i.e., RQ3, remained unanswered in [20]. However, we identified [21] as a work from traditional process mining that contributes to the confidentiality of the results, i.e., RQ3. Therefore, this is a potential starting point for discussing RQ3 at the example of directly-follows graphs.

Currently, the evaluation of federated process mining research, poses a challenge itself. As federated process mining is currently emerging, setting up a quantitative approach remains complicated as, to the best of our knowledge, no public, inter-organizational event data exists. Therefore, qualitative arguments on complexity and on correctness are necessary. A quantitative analysis is currently only possible by creating artificial event logs, e.g., by splitting existing event logs to be partial event logs of several organizations.

For the implementation of the process mining algorithms, we consider to continue on designing multi-party computation protocols that exploit homomorphic encryption as already exhibited by [20]. This is due to the fact, that it is considered as a suitable technology to meet the demands of privacy regulation laws and since it allows to design protocols that match the arithmetic used in the algorithms of our traditional process mining methods; thus, allowing for verifiable federated process mining algorithms that reproduce the results from traditional process mining. In detail, we expect fully homomorphic encryption to allow for the computations on the encrypted data without the necessity to share the data in between. For the tooling and since we aim to optimize performance of our algorithms, we identified the Rust4PM [22] and TFHE-rs [23] libraries as performant Rust implementations to build up on for the use of process mining and fully homomorphic encryption, respectively.

For our publication strategy, we expect to write several papers that consider RQ1 and RQ3 as we want to implement a minimal federated process mining toolkit that covers process discovery, conformance checking, and process enhancement. Since RQ2 influences the speed up of the designed algorithms, we plan to dedicate a single paper to this matter. Finally, we focus on RQ4 to do a user study in which we highlight the use and the results of the previously designed algorithms.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

[1] W. M. P. van der Aalst, Intra- and inter-organizational process mining: Discovering processes within and between organizations, in: PoEM, volume 92 of *Lecture Notes in Business Information Processing*, Springer, 2011, pp. 1–11.

[2] Health insurance portability and accountability act of 1996, Public law 104 (1996) 191.

[3] Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free

movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance), 2016.
[4] F. Mannhardt, Responsible process mining, in: Process Mining Handbook, volume 448 of *Lecture Notes in Business Information Processing*, Springer, 2022, pp. 373–401.
[5] G. Elkoumy, S. A. Fahrenkrog-Petersen, M. F. Sani, A. Koschmider, F. Mannhardt, S. N. von Voigt, M. Rafiei, L. von Waldthausen, Privacy and confidentiality in process mining: Threats and research challenges, ACM Trans. Manag. Inf. Syst. 13 (2022) 11:1–11:17.
[6] W. M. P. van der Aalst, Federated process mining: Exploiting event data across organizational boundaries, in: SMDS, IEEE, 2021, pp. 1–7.
[7] M. Rafiei, W. M. P. van der Aalst, An abstraction-based approach for privacy-aware federated process mining, IEEE Access 11 (2023) 33697–33714.
[8] R. Gatta, M. Vallati, J. Lenkowicz, C. Masciocchi, F. Cellini, L. Boldrini, C. Fernández-Llatas, V. Valentini, A. Damiani, On the feasibility of distributed process mining in healthcare, in: ICCS (5), volume 11540 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 445–452.
[9] L. Nucciarelli, R. Gatta, A. M. Tudor, E. Tavazzi, G. Arcuri, M. Vallati, G. Ibáñez-Sánchez, Z. Valero-Ramon, C. Fernández-Llatas, A. Damiani, Towards distributed process discovery in healthcare: Testing and proving the feasibility of the federated alpha+ algorithm, in: AIME (2), volume 15735 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 294–299.
[10] J. D. Hernandez-Resendiz, E. Tello-Leal, H. M. Marin-Castro, U. M. Ramirez-Alcocer, J. A. Mata-Torres, Merging Event Logs for Inter-organizational Process Mining, Springer, 2021, pp. 3–26.
[11] M. Rafiei, M. Pourbafrani, W. M. P. van der Aalst, Federated conformance checking, Inf. Syst. 131 (2025) 102525.
[12] H. Zhian, R. Buyya, A. Polyvyanyy, Federated stochastic process discovery using grammatical inference, in: CAiSE (2), volume 15702 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 76–93.
[13] A. Acar, H. Aksu, A. S. Uluagac, M. Conti, A survey on homomorphic encryption schemes: Theory and implementation, ACM Comput. Surv. 51 (2018) 79:1–79:35.
[14] D. Evans, V. Kolesnikov, M. Rosulek, A pragmatic introduction to secure multi-party computation, Found. Trends Priv. Secur. 2 (2018) 70–246.
[15] G. Elkoumy, S. A. Fahrenkrog-Petersen, M. Dumas, P. Laud, A. Pankova, M. Weidlich, Secure multi-party computation for inter-organizational process mining, in: BPMDS/EMMSAD@CAiSE, volume 387 of *Lecture Notes in Business Information Processing*, Springer, 2020, pp. 166–181.
[16] V. Goretti, D. Basile, L. Barbaro, C. D. Ciccio, Trusted execution environment for decentralized process mining, in: CAiSE, volume 14663 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 509–527.
[17] S. J. J. Leemans, D. Fahland, W. M. P. van der Aalst, Discovering block-structured process models from event logs - A constructive approach, in: Petri Nets, volume 7927 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 311–329.
[18] J. Rott, M. Böhm, H. Krcmar, Laying the ground for future cross-organizational process mining research and application: a literature review, Bus. Process. Manag. J. 30 (2024) 144–206.
[19] I. Ileri, T. G. Erdogan, A. K. Tarhan, Privacy for process mining: A systematic literature review, IEEE Access 13 (2025) 83171–83194.
[20] C. Rennert, J. Albers, S. J. J. Leemans, W. van der Aalst, Your secret is safe with me: Federated directly-follows graph discovery, in: ICPM, IEEE, 2025.
[21] S. A. Fahrenkrog-Petersen, M. Kabierski, H. van der Aa, M. Weidlich, Semantics-aware mechanisms for control-flow anonymization in process mining, Inf. Syst. 114 (2023) 102169.
[22] A. Küsters, W. M. P. van der Aalst, Rust4pm: A versatile process mining library for when performance matters, in: BPM (Demos / Resources Forum), volume 3758 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 91–95.
[23] L. Brenna, I. S. Singh, H. D. Johansen, D. Johansen, Tfhe-rs: A library for safe and secure remote computing using fully homomorphic encryption and trusted execution environments, Array 13 (2022) 100118.