Process Mining in Trusted Execution Environments: Towards Hardware Guarantees for Trust-Aware Inter-organizational Process Analysis

Marcel Müller[1,2(✉)], Anthony Simonet-Boulogne[3], Souvik Sengupta[3], and Oliver Beige[2]

[1] Technische Universität Berlin, Berlin, Germany
marcel.mueller@tu-berlin.de
[2] JadenX Research, Berlin, Germany
{marcel.mueller,oliver.beige}@jadenx.com
[3] iExec Blockchain Tech, Lyon, France
{anthony,souvik}@iex.ec

Abstract. Process mining techniques enable business process analysis on event logs extracted from information systems. Currently, industry applications and research in process mining predominantly analyze intra-organizational processes. Intra-organizational processes deal with the workflows within a single organization. However, analyzing inter-organizational processes across separate companies has the potential to generate further insights. Process analysts can use these insights for optimizations such as workflow improvements and process cost reductions. It is characteristic for inter-organization process analysis that it is not possible to uncover the insights by analyzing the event logs of a single organization in isolation. On the other hand, privacy and trust issues are a considerable obstacle to adopting inter-organizational process mining applications. The independent companies fear competitive disadvantages by letting third parties access their valuable process logs. This paper proposes a concept for inter-organizational process mining using trusted execution environments in a decentralized cloud. The hardware-based approach aims to technically prevent data leakage to unauthorized parties without the need for a trusted intermediary. The contributions of this paper are theoretical and identify future research challenges for implementing the concept.

Keywords: Process Mining · Privacy · Trusted Execution Environments · Inter-organizational Process Mining

1 Introduction

Process mining analyzes the real-world execution of business processes. The analysis utilizes event logs extracted from information systems to construct a

J. Munoz-Gama and X. Lu (Eds.): ICPM 2021 Workshops, LNBIP 433, pp. 369–381, 2022.
https://doi.org/10.1007/978-3-030-98581-3_27

business process model [1]. A variety of process mining techniques and configurations enable analysts to derive different insights into their processes. Process mining can be used to identify compliance violations, find process bottlenecks, and investigate the root causes of undesired process behavior. Usually, process mining analyzes processes within a specific organization (intra-organizational processes). Yet, in practice, inter-organizational workflows are standard in various industries. In an inter-organizational business process, different organizations execute separate parts of a shared workflow. Examples of such processes include e-commerce, supply chain management, or international bank transactions. However, analyzing data from other companies is trust- and privacy-intensive [2–4]. Event logs record valuable information of an organization's real-life operation details. These details can be exploited to analyze a collaborator's internal processes and to gain competitive advantages. Thus, many organizations refrain from participating in inter-organizational process mining and optimization. However, mining inter-organizational processes as a whole can enable different insights and benefits that cannot be derived from analyzing the private processes of collaborators in isolation. All parties may benefit from such a high-level analysis.

Figure 1 illustrates an example inter-organizational hiring processes. The process model shows how a certain company (the *seeking company*) finds new employees for its software development jobs. Since recruiting processes are time- and cost-intensive, the company outsources the initial recruiting task to three independent recruiters. The process starts with the seeking company defining the job requirements. Afterward, the organization contacts three different recruiters in parallel for the first-level candidate screening. Their task is to find the best-suited candidates for their job opening. The three recruiters have different strategies to find the best candidates. Recruiter 1 approaches the task by searching candidates on professional platforms like LinkedIn[1]. The recruiter sends cold messages to candidates and conducts a general pre-interview with them. The objective of the pre-interview is to find out if all formal requirements for the candidate to become a potential employee are fulfilled. This might include, for instance, having the right working permits. Afterward, Recruiter 1 conducts a technical interview to see if the candidate has the right skill set for the job. In the end, Recruiter 1 decides whether the candidate is suited for the position. If yes, the recruiter forwards the CV to the seeking company. Recruiter 2 has a different approach. This recruiter makes a job post on an open online job board like Indeed[2]. After a while, the recruiter receives some applications and assesses the CVs of the candidates. Recruiter 2 does not conduct general interviews and proceeds directly to the technical interview. After the technical interview, this recruiter also decides and forwards the candidate to the seeking company. Recruiter 3 starts the candidate search on a professional platform, like Recruiter 1. However, Recruiter 3 is not a technical expert and does not conduct technical interviews after the general pre-interview. All three recruiters forward

[1] https://www.linkedin.com/.
[2] https://de.indeed.com/.

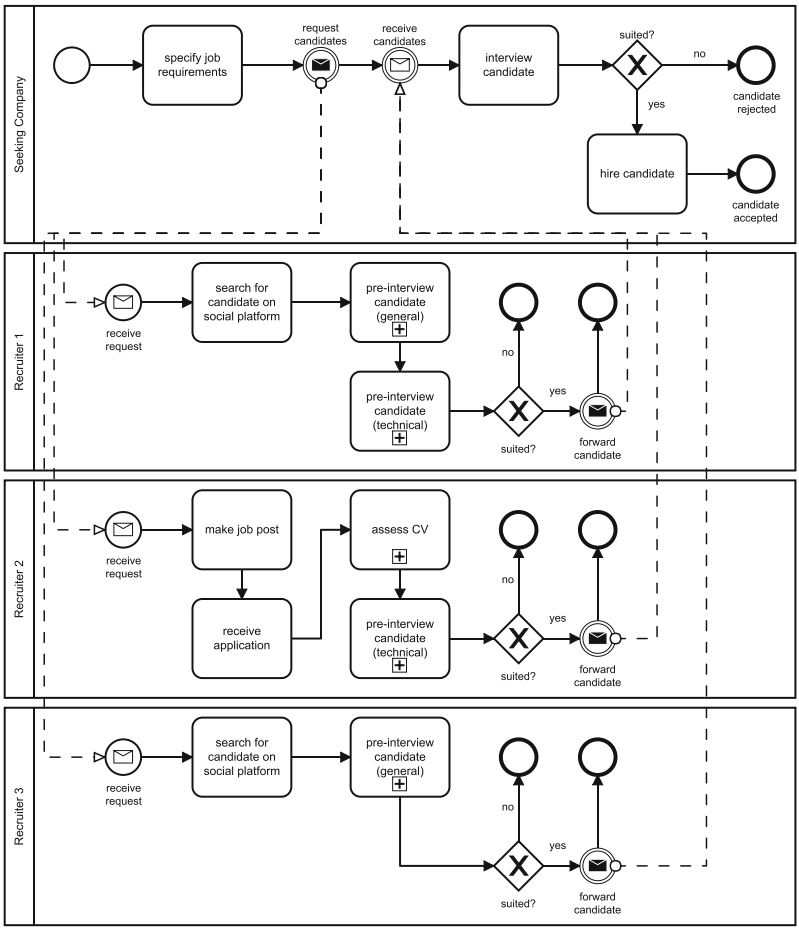

Fig. 1. Example inter-organizational process in human resources using the BPMN 2.0 standard [5].

their candidates to the seeking company, where a final interview is conducted. Afterward, the decision of whether or not to hire the candidate is made.

This example shows an inter-organizational process where all four organizations act independently. Yet, they have a common goal to make to recruiting process as efficient as possible. Especially the seeking company wants to reduce their recruiting time. Their final interviews are conducted by the seeking company's most skilled tech specialists. This circumstance makes every final interview cost-intensive. In such a case, gathering the event logs from all recruiters and applying process mining techniques can help determine process dependencies and causalities. However, recruiters do not want to disclose details of their recruiting process to third parties. They experimented in the past to find the

best recruiting strategy. Thus, by sharing their detailed approaches, they could have competitive disadvantages.

This paper proposes a theoretical concept of executing process mining tasks in trusted execution environments (TEEs). TEEs are hardware-based approaches that enable processing data in a secure enclave. Thus, there is no *technical* possibility to leak the information to unauthorized third parties. The outcome is a theoretical concept and an identification of research challenges that need to be solved before the concept can be implemented in practice.

The remainder of this paper is structured as follows. Section 2 reviews the current state of the art in privacy-preserving inter-organizational process mining and trusted computing. Afterward, Sect. 3 introduces our novel concept for privacy-aware process mining in TEEs. Section 4 discusses the implementation challenges of the concept, before Sect. 5 concludes on the impact of this scientific contribution.

2 Related Work

Recently, privacy aspects of inter-organizational process mining have seen an increase in academic and professional interest. The current state of the art in privacy-preserving process mining approaches can be divided into two main segments. The first segment focuses on privacy preservation of information related to a *individuals* encoded in a process log, e.g., employee information. The other group of approaches focuses on protecting the information of an *organization* and its business secrets.

Individual-focused privacy-preserving process mining approaches focus on the privacy of the information of individuals that are included in event logs. These concepts have applications in fields like individual health care or manufacturing workflows [6]. For instance, a large hospital might want to analyze its emergency room response processes. Therefore, they need data related to specific cases of emergency room arrivals. The event log might include information specific to a patient. Regulatory frameworks like the General Data Protection Regulation (GDPR) [7] or the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [8] require this data to be protected. Current research on individual privacy-preserving process mining employs concepts such as differential privacy and k-anonymity [9–11]. In addition, other approaches employ encryption as the main method of privacy preservation. These concepts use standards like the Advanced Encryption Standard (AES) [12] or the Pallier Cryptosystem [13] to encrypt personal information encoded in the event log. Such cryptography-based concepts often also enable privacy preservation of business-related information as well [14,15].

Organization-focused privacy-preserving process mining techniques have the ultimate goal to protect business-related information. The leakage of valuable organizational information to third parties might lead to compliance issues and competitive disadvantages. Thus, especially inter-organizational business processes pose a specific challenge to such organization-focused privacy-preserving

process mining techniques. It is characteristic of inter-organizational process mining to acquire event logs from different organizations. These joint insights may generate a greater value than the insights created from separate event logs in an isolated manner. Current research in this field proposed different concepts for executing the mining process. One approach is to mine partitioned data in a decentralized fashion. The partitions can be used to generate a common model without revealing the raw data [16]. The common model may differ for organizations since they all have different knowledge. Cryptography-based approaches encrypt the valuable information [14]. Lately, also approaches that use private computing paradigms such as homomorphic encryption [15] or secure multi-party computation [17] emerged. Both of these approaches are famously known as the privacy-preserving computation (PPC) methodologies. Lately, trusted execution environments (TEEs) have been introduced [18]. TEEs are a hardware-level privacy-aware computation paradigm. This paper presents a theoretical concept for organization-focused privacy-preserving process mining using TEEs.

Trusted execution environments (TEEs) are a hardware-based approach for trusted computation provided by some modern micro-processors, e.g. Intel *Software Guard Extension* (SGX) [18] and ARM Trustzone [19]. The main component of a TEE is a secure element that resides within a separate area of the CPU chip. Code and data in the secure element are entirely isolated from other programs and from the host operating system. This paradigm protects the data from theft and the code from tampering. TEEs, and Intel SGX in particular, provide low-level primitives for defining specific rules (e.g. which software package can decrypt a dataset). These rules are enforced by using hardware-based cryptography. However, expressing complex multi-processor workflows like the one we have described above requires a higher-level rule system. This rule system is in charge of orchestrating the encryption of the input data, the provisioning of several secure enclaves, and the dataflow between them. Distributed ledger technologies (DLTs) offer a decentralized execution environment with an immutable record of transactions. Thus, DLTs can be a well-suited platform for orchestrating process mining. The organizations can use smart contracts to define authorizations, to record job requests, and to verify remote attestations (proof of a correct execution in TEE) with no risk of their intent being altered. Current approaches combining distributed ledgers (for expressing rules) and TEEs (for enforcing them) for trusted computing include iExec [20] and Ekiden [21].

The primary purpose for adopting the hardware-based TEE (HW TEE) in our contribution unfolds as follows. Besides enabling data integrity and confidentiality, HW TEEs also ensure code integrity, code confidentiality, programmability, attestability, recoverability, and authenticated application launch facilities. Thus, these characteristics make the HW TEE a suited option for the organizations for doing privacy-preserving process mining.

3 A Concept for Privacy-Aware Process Mining in Trusted Execution Environments

In this paper, we introduce a theoretical concept on how TEEs and blockchains can be used for privacy-aware inter-organizational process mining. The setup of the orchestration of the TEEs is inspired by the iExec decentralized cloud computing framework[3]. However, the general concepts presented in the following are independent of any framework to orchestrate TEEs.

3.1 System Architecture

Our concept consists of different system components. The following paragraphs introduce them and their workflows in the inter-organizational mining process. The architecture diagram in Fig. 2 visualizes the interactions.

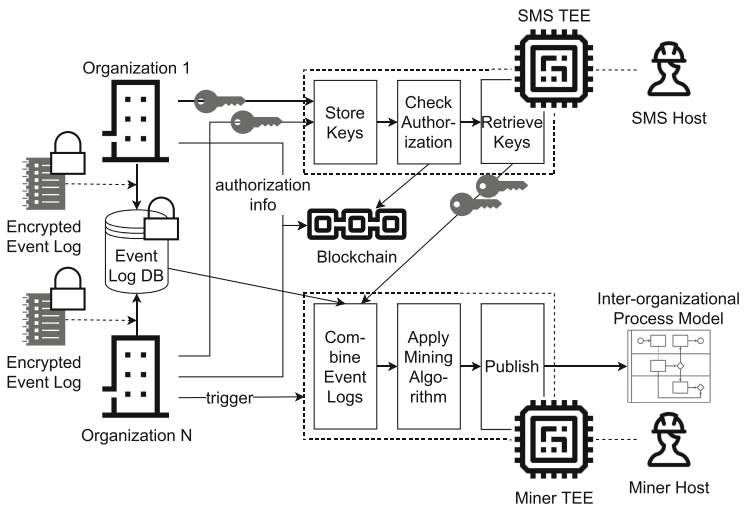

Fig. 2. Architecture diagrams of different roles interacting with each other.

Organizations. In our concept, N organizations have their private information system where they acquire new events. These private information systems are isolated from each other.

Secret Management System. The secret management system (SMS) is a key component that acts as a secure intermediary between the organizations that provide data and the process miner which processes it. Because TEEs require

[3] https://iex.ec.

data be encrypted specifically for a given enclave session, the workflow could not run asynchronously without it. In our example, this would translate to having the organizations encrypt their logs after the process mining has started. In our design, the SMS itself is a secure application running in an enclave. It receives the decryption keys of all the data from the organizations and manages it according to the authorized orders. The orders are signed and recorded on the blockchain. The SMS is thus a critical component that holds all of the decryption keys to every log file, but the fact that it runs in a TEE guarantees that only no one and nothing besides its code can access them, not even the administrator of the machine it is running on.

Blockchain. In our concept, we use the blockchain as tamper-proof storage of authorization statements using smart contracts [22]. The SMS is only allowed to give the keys to authorized entities. Thus, the organizations create transactions to trigger smart contracts stating which miner can retrieve the keys for a specific order.

Event Log Database. The public event database stores the encrypted event log files of all organizations. This ensures ensure the integrity of the private logs and enables the inter-organizational mining process to retrieve them. The database host never has access to the keys of the event logs.

Miner. The miner is responsible for applying process mining techniques to the combined event logs of different organizations. The three-step subprocess consists of combining the event logs, mining the process models, and making the insights available to the organizations. All mining tasks are executed in a TEE so that the miner host cannot interfere.

3.2 Workflow

The following sections describe the process of privacy-preserving process mining using TEEs in detail. Figure 3 shows the process model of our concept.

Prerequisites. We assume that the following activities happened before the core process. All N organizations need to synchronize the case ids and select a process mining technique upfront. In process mining, a *case* is a unique identifier that groups a set of events. All events in a case belong logically together. In the running example, every instance of a recruiting process of a backend developer consists of different events. Such events may be that a recruiter found a new candidate or that the hiring company made a decision. All events that belong to the same *instance* of the process can be grouped together in a case. A sequential order of timestamped events within a case is called a *trace*. The organizations might use different systems to track the events that fall into their domain. Thus, they need to synchronize case identifiers. In that way, the process miner can later merge different sub-traces that belong to the same case in a TEE. It is also

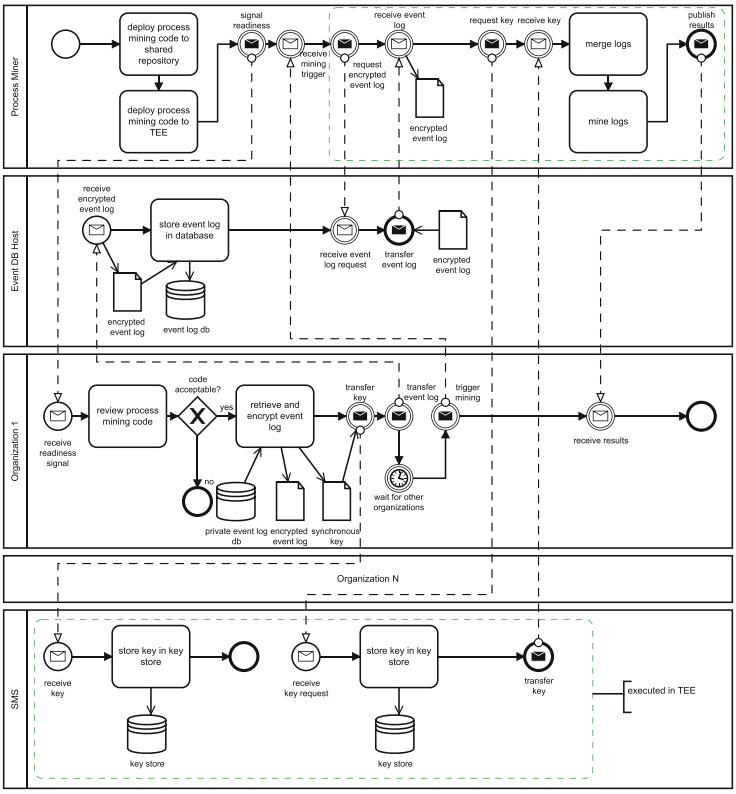

Fig. 3. Illustration of the high-level process of privacy-preserving process mining using TEEs. The green boxes indicate workflow executed within a TEE. The diagram uses the BPMN 2.0 standard for illustration [5]. The collapsed pool indicates a variable number of organizations that all follow the same logic. (Color figure online)

required that the process miner is in possession of an implementation of the process mining technique. Finally, each organization generates a unique symmetric encryption key and uses it to encrypt an archive containing all of their log files.

Initialization. The core workflow component for privacy-aware process mining in TEEs is the process mining code, which must be audited and approve it by all of the organizations. In practice, the source code is shared in a repository that all organizations can access. All organizations audit the code and if approved they record a rule linking the hash value of the packaged code to the hashed value of the encrypted archive. This requirement ensures that the SMS will later give the right program access to the right data. In our concept, the rules are recorded in a blockchain smart contract [23] to preserve the integrity of the hashes in a decentralized fashion. All organizations have blockchain peers and guarantee the integrity of transactions.

Review and Data Contribution. After the deployment phase, the process miner signals readiness to all N organizations. The organizations can review the process mining code. In this review, they can assess if the code meets all privacy and compliance requirements. In case an organization decides that the code does not fulfill its requirements, the process terminates. After a positive code review, the organizations retrieve their event logs from their private information systems. They encrypt their event logs separately. Therefore, an organization needs to encrypt its event log using a symmetric encryption mechanism, such as AES-256 [12]. A hash of the encrypted data set is submitted to the blockchain with a transaction. The organizations can later use the hash to ensure the integrity of the data set. The encrypted event log itself is stored at an independent host as an encrypted file. The organizations share their symmetric key with a secret management service (SMS) of their choice. This SMS is a simple program that lets only authorized entities access keys. The right functionality of the SMS can be guaranteed because the SMS program is also executed in a TEE. Its code is open source. Thus, every entity can audit its code. Through the attestation that a TEE produces, it is possible to prove that only the desired program has been executed and nothing else.

Mining. A organization needs to trigger the process mining task. This trigger is expressed through a blockchain transaction as an *execution order* that needs to be signed by the requesting organization. The execution order specifies which code (the selected process mining technique) should be executed and which input data sets (the N event logs of the organizations) should be mined. The process miner starts the mining process in a TEE. In that way, the host does not have any influence on the execution of the mining program, and a remote attestation proves the correct execution. First, the process miner requests all the encrypted event logs from the independent event data storage; then, it makes a request to the SMS to obtain the keys to decrypt the process logs. The SMS uses TEE primitives to verify that the miner is actually running in an enclave. The SMS only allows it to access the keys if there is an order that assigns the miner to a task that includes the data sets of the respective organizations. The correctness of this logic can be guaranteed through the attestation of the TEE. After the process miner received and decrypted all event logs in the enclave, the merging of the logs begins. While merging, the different sub-traces of the separate organizations are used to end-to-end traces. This trace reflects the full inter-organizational processes with all the sub-processes of the collaborators. The merging process yields a full event log that the selected process mining technique can then mine.

Insights. In the end, the process mining TEE compiles the aggregated result of the mining process. These result is a joined process model that encompasses the whole inter-organizational process. Furthermore, the TEE distributes them to all N organizations. In that way, the N organizations only get insights from the merged and aggregated process. However, they cannot get any insights into the sub-processes of a specific organization.

3.3 Security Comparison

The following sections discuss the security and privacy features of our proposed approach to currently existing technical foundations for privacy-preserving process mining. Secure multiparty computation (SMPC), homomorphic encryption (HE), and differential privacy (DP) enable different mechanisms for ensuring privacy and confidentiality in inter-organizational process mining. SMPC keeps the executing system from exposing the input data [24]. However, it does not provide any guarantees for the output data. HE mitigates potential vulnerabilities in the storage or computing environment from compromising the data. However, in the case of HE, if some party gets the access privilege, the authorized party can easily access entire datasets [25]. DP provides a layer of privacy by obfuscation in case some data concerning individual entities leaks. Yet, it can not counteract vulnerabilities in the infrastructure used to store or manage the data [25]. Thus, executing process mining code in a hardware TEE differs from the current concepts for privacy-preserving process mining. It ensures complete computation confidentiality through memory encryption at the hardware level. Inputs and outputs to computing tasks are encrypted. This makes hardware-based TEEs suitable for developing our multi-organizational trusted process mining framework, as long as the user trusts the hardware design.

4 Implementation Challenges

To implement our presented concept, we need an orchestration layer that can provide provision TEE resources for process mining tasks. Therefore, we adopt a *decentralized cloud* paradigm. There, workers can contribute their TEE resources to a process mining task. We adopt this paradigm so that the organizations do not have to deal with the overhead of setting up TEE resources on their premises. Furthermore, the incentivization mechanism ensures that attestations are always distributed to all involved organizations.

Currently, the iExec framework [20] and Hyperledger Avalon [26] are the two only decentralized cloud computing frameworks that can orchestrate Intel SGX enclaves [18]. Both utilize the blockchain to store orders, resource allocations, and attestation securely. In the following, we explore the steps needed to implement the presented concepts using iExec. We make this choice because it is more advanced in its development maturity than other approaches. However, our principles are independent of any framework.

The iExec worker infrastructure is deployed on top of the Ethereum blockchain. A suite of support tools allows anyone to record TEE applications packaged as Docker containers. Data management tools enable the management of encrypted data sets and setting fine-grained authorization rules. These authorizations include which application can access which data set and which users can trigger an execution. The authorization is implemented in iExec with a secret management service (SMS) similar to our proposed inter-organizational process mining concept.

While the iExec framework is the most suited candidate to implement our concepts, further development is required to support some specific needs for process mining. Namely, this includes the possibility of assigning several data sets to a single execution and distributing the result to multiple users. At the time of writing, iExec does not support consuming multiple data sets in a single execution. Furthermore, the SMS in iExec is in a prototype stage. Its full implementation in an Intel SGX TEE is still not complete. The upcoming release of SGX 2 CPUs by Intel should significantly improve the performance and scalability of the service.

Once these challenges are overcome, the novel approach to privacy-preserving inter-organizational process mining as presented in this paper can be researched, implemented, and evaluated further.

5 Conclusion

In this paper, we introduced a novel concept for privacy-aware inter-organizational process mining using trusted execution environments. The contributions are theoretical. We identified challenges for future work that need to be solved to implement the concept.

Our concept can enable process mining in application domains with sensitive data that currently do not utilize process analysis in cross-organizational processes. The first application area we foresee is supply chain management. Several logistics companies must collaborate to transport a parcel from a sender to a receiver in international deliveries. All companies want to optimize their workflows as much as possible. The inclusion of the whole inter-organizational process could help optimize shipping times and improve customer satisfaction. Another application area is fraud in finance. Currently, detecting money laundry circles requires transaction logs from different banks. Due to the privacy requirements of their customers, banks are reluctant to share data with any third party. Introducing our concept for money laundry detection could build trust since all processing steps of shared data can be audited, and the TEEs guarantee that no other unauthorized code is executed.

References

1. van der Aalst, W.: Process Mining: Data Science in Action. Springer, Berlin (2016). https://doi.org/10.1007/978-3-662-49851-4
2. Müller, M., Ostern, N., Koljada, D., Grunert, K., Rosemann, M., Küpper, A.: Trust mining: analyzing trust in collaborative business processes. IEEE Access **9**, 65044–65065 (2021)
3. Müller, M., Garzon, S.R., Rosemann, M., Küupper, A.: Towards trust-aware collaborative business processes: an approach to identify uncertainty. IEEE Internet Comput. **24**(6), 17–25 (2020)
4. Elkoumy, G.: Privacy and confidentiality in process mining-threats and research challenges (2021). arXiv preprint: arXiv:2106.00388

5. OMG. Business process model and notation (BPMN), version 2.0. https://www.omg.org/spec/BPMN/2.0/PDF. Accessed on 29 July 2021
6. Mannhardt, F., Petersen, S.A., Oliveira, M.F.: Privacy challenges for process mining in human-centered industrial environments. In: 2018 14th International Conference on Intelligent Environments (IE), pp. 64–71. IEEE (2018)
7. Directive 95/46/ec (general data protection regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed 30 July 2021
8. Health insurance portability and accountability act of 1996 public law 104–191 (1996). https://www.govinfo.gov/content/pkg/PLAW-104publ191/html/PLAW-104publ191.htm. Accessed 30 July 2021
9. Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M., Michael, J.: Privacy-preserving process mining. Bus. Inf. Syst. Eng. **61**(5), 595–614 (2019)
10. Fahrenkrog-Petersen, S.A., van der Aa, H., Weidlich, M.: PRETSA: event log sanitization for privacy-aware process discovery. In: 2019 International Conference on Process Mining (ICPM), pp. 1–8. IEEE (2019)
11. Fahrenkrog-Petersen, S.A., van der Aa, H., Weidlich, M.: PRIPEL: privacy-preserving event log publishing including contextual information. In: Fahland, D., Ghidini, C., Becker, J., Dumas, M. (eds.) BPM 2020. LNCS, vol. 12168, pp. 111–128. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58666-9_7
12. Daemen, J., Rijmen, V.: AES proposal: Rijndael (1999)
13. Paillier, P.: Public-key cryptosystems based on composite degree residuosity classes. In: Stern, J. (ed.) EUROCRYPT 1999. LNCS, vol. 1592, pp. 223–238. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48910-X_16
14. Burattin, A., Conti, M., Turato, D.: Toward an anonymous process mining. In: 2015 3rd International Conference on Future Internet of Things and Cloud, pp. 58–63. IEEE (2015)
15. Tillem, G., Erkin, Z., Lagendijk, R.L.: Mining encrypted software logs using alpha algorithm. In: SECRYPT, pp. 267–274 (2017)
16. Liu, C., Duan, H., Zeng, Q., Zhou, M., Faming, L., Cheng, J.: Towards comprehensive support for privacy preservation cross-organization business process mining. IEEE Trans. Serv. Comput. **12**(4), 639–653 (2016)
17. Elkoumy, G., et al.: Secure multi-party computation for inter-organizational process mining. In: Nurcan, S., Reinhartz-Berger, I., Soffer, P., Zdravkovic, J. (eds.) BPMDS/EMMSAD -2020. LNBIP, vol. 387, pp. 166–181. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49418-6_11
18. Costan, V., Devadas, S.: Intel SGX explained. IACR Cryptol. ePrint Arch. **2016**(86), 1–118 (2016)
19. Pinto, S., Santos, N.: Demystifying ARM TrustZone: a comprehensive survey. ACM Comput. Surv. (CSUR) **51**(6), 1–36 (2019)
20. Zhang, L., Bakshi, S., Zao, K.: Off-chain trusted computing. IEEE Internet Things Mag. **3**(2), 8–9 (2020)
21. Cheng, R., et al.: Ekiden: a platform for confidentiality-preserving, trustworthy, and performant smart contracts. In: 2019 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 185–200. IEEE (2019)
22. Müller, M., Ostern, N., Rosemann, M.: Silver bullet for all trust issues? Blockchain-based trust patterns for collaborative business processes. In: Asatiani, A., et al. (eds.) BPM 2020. LNBIP, vol. 393, pp. 3–18. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58779-6_1
23. Buterin, V., et al.: Ethereum white paper. GitHub Repos. **1**, 22–23 (2013)

24. Sayyad, S.: Privacy preserving deep learning using secure multiparty computation. In: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 139–142. IEEE (2020)
25. Zorarpacl, E., Ozel, S.A.: A hybrid approach of homomorphic encryption and differential privacy for privacy preserving classification. Int. J. Appl. Math. Electron. Comput. **8**(4), 138–147 (2020)
26. Hyperledger avalon. https://github.com/hyperledger/avalon. Accessed 30 Aug 2021