Inter-organizational Process Mining through Trusted Execution Environments

Davide Basile[1][0000−1111−2222−3333], Luca Barbaro[1][1111−2222−3333−4444], Valerio Goretti[1][0000−0001−9714−4278], and Claudio Di Ciccio[1][2222−−3333−4444−5555]

Sapienza University of Rome

Abstract. ...

1 Introduction

In today's interconnected business landscape, organizations are constantly seeking ways to enhance their operational efficiency, increase their performance, and gain valuable insights to improve their processes. In this context, the availability of worthwhile information plays a key role. One of the primary obstacles lies in securely and reliably accessing and utilizing data from various companies or business units, ensuring that all involved parties can derive substantial benefits from it. Traditional approaches to sharing sensitive data across organizational boundaries often involve concerns related to privacy, data integrity, and the need for mutual trust. This paper introduces a novel approach that leverages Trusted Execution Environment (TEE) technology to facilitate the exchange of event logs among different companies or business units. TEE provides a secure and isolated environment within a computer system, ensuring the confidentiality, integrity, and privacy of data and code execution. By utilizing TEE, organizations can exchange event logs in a trusted and privacy-preserving manner, enabling them to harness valuable information from external processes to enhance, monitor, or modify their own processes effectively.

The proposed methodology combines the power of TEE with process mining techniques to unlock the potential of inter-organizational event log exchange. Process mining is a data-driven approach that extracts knowledge and insights from event logs to gain a comprehensive understanding of business processes. By incorporating external event logs from trusted sources, organizations can obtain a broader and more accurate view of the end-to-end processes they are involved in, leading to better process optimization, performance monitoring, and decision-making. By bridging the gap between different organizations, the utilization of TEE-enabled event log exchange for process mining offers promising opportunities for collaborative learning, benchmarking, and continuous process improvement. The findings and insights derived from this research have the potential to revolutionize the way organizations approach process optimization and enable them to make more informed decisions based on a holistic view of their processes.

2 Related Work

The literature proposes several studies that consider process mining techniques in inter-organizational environments. Van Der Aalst [1] shows that inter-organizational processes can be divided according to different dimensions making identifiable challenges of inter-organizational process extractions. Elkoumy et al. [2] propose a tool that allows independent parts of an organization to perform process mining operations by revealing only the result. This tool is called Shareprom and exploits the features of secure multi-party computation (MPC). Engel et al. [3] present EDImine Framework, which allows to apply process mining operations for inter-organizational processes supported by the EDI standard[1] and evaluate their performance using business information. Elkoumy et al. [4] propose an architecture based on MPC. This architecture aims to perform process mining operations without sharing their data or trusting third parties.

Applying process mining techniques in intra-organizational contexts requires merging the event logs of the organizations participating in the process. Hernandez-Resendiz et al. [5] present a methodology for merging logs at the trace and activity level using rules and methods to discover the process. [6] [7] [8] [9] [10]

Once the data exchange has taken place, it is critical that the data be stored in a trusted part of the consumer's device. Basile et al. [11] in their study created a framework called Regov that allows for the exchange of sensitive information in a decentralized web context, ensuring usage control-based data access and usage. To ensure control over the consumer's device, Davide et al. use trusted execution environment that allows storage and utilization management of retrieved resources.

The application of process mining in an inter-organizational scenario is infrequent due to concerns related to privacy and confidentiality, integrity and data heterogeneity. To overcome these problems, a large number of techniques have been proposed. Federated process mining [12] aims to effectively manage the problem of privacy-preserving. Using federated data sources, event data can be transparently mapped between multi-source autonomous provider to monitor, analyze, and improve processes across organizations.

3 Motivating Scenario

4 High level Architecture

In the following section we present the high level architecture underlying our solution. Therefore, we take into account each component individually. Subsequently, we provide an overview on the main interactions taking place between the introduced components.

[1] https://edicomgroup.com/learning-center/edi/standards

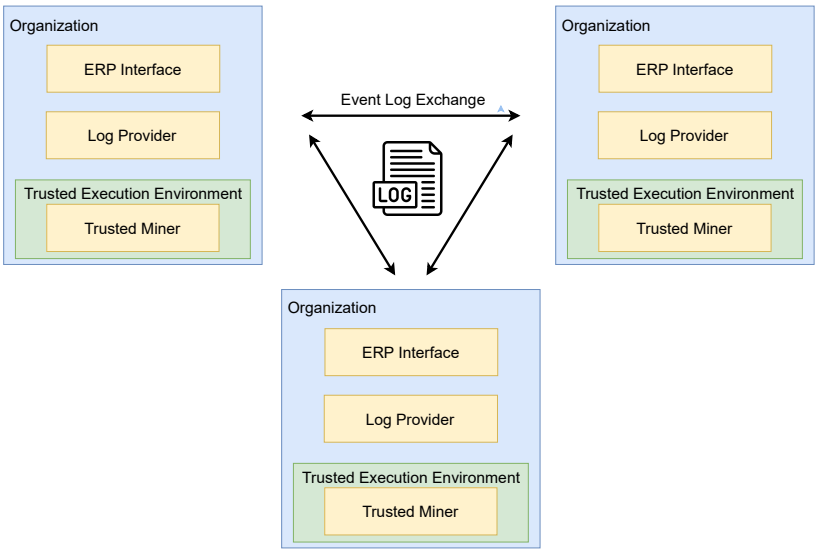

Fig. 1. High-level architectural overview of the framework.

4.1 Main components

Our architecture involve networks of nodes controlled by different `Organizations` exchanging their event logs. `Organizations` in the same network collaborate to reach a common objective composing business processes whose event logs are defraged over multiple places. The hospital and the specialized clinic, mentioned in the running example, provide an example of partner organizations. An `Organization` assumes simultaneously two different roles. We refer to an `Organization` as a *miner* when it applies process mining algorithms using local event logs in combination with ones generated by partner `Organizations`. When an `Organization` deliver event logs to be mined, we call it *provider*. Each `Organization` is associated with an asymetric key couple thanks to which it authenticates and makes authenticable messages from/to other `Organizations`.

In section 4, we propose an high level schematization of our solution. Each `Organization` embeds three main components: the `ERP Interface`, the `Log Provider` and the `Trusted Miner`. In the next paragraph, we address the newly mentioned components.

ERP Interface The `ERP Interface` collects the logic to interact with the Enterprise Resource Planning (ERP) system of the `Organization`. ERP systems helps `Organizations` to handle business processes including accounting and resource management. The mantainance of event logs is one of the many tasks performed by these systems. In our architecture, we generalize the interaction with these systems through the `ERP Interface`. The `ERP Interface` provides

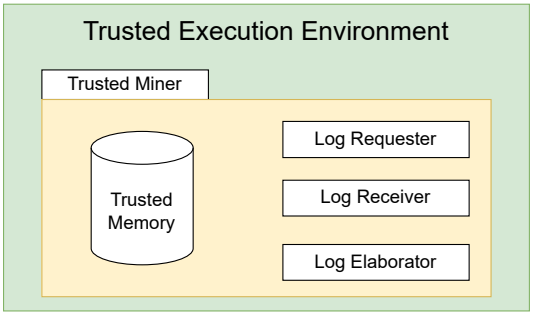

Fig. 2. Modules of the Trusted Miner component running inside Trusted Execution Environments

the local `Log Provider` and `Trusted Miner` access to event logs generated inside the `Organization`.

Log Provider The `Log Provider` component delivers on-demand data to `Trusted Miner`s belonging to partner `Organization`s. It handles event log request through access control methodologies, thanks to which the identity of the miner is determined and the access to the resource is decided.

`Log Provider`s authenticate event log requests using asymetric encryption methodologies. Through the latter, `Log Provider`s verifiy parameters embedded in the data request. The goal of the authentication procedure is to extract the public key representing the identity of the sender `Trusted Miner`. In order to deliver data, the `Log Provider` retrieve the local event log communicating with the `ERP Interface` of the `Organization`.

Remote Attestation and Log Segmentation are crucial procedures handled by the `Log Provider` during the provision process. Through Remote Attestation, `Log Provider`s verify that the `Trusted Miner` that has generated the log request is: (i) a known software object running inside a trusted execution environment; (ii) controlled by a partner `Organization` that has rights to access the event log. We named Log Segementation the process through which `Log Provider`s split the event log to be delivered in sub-log of smaller size.

Our architecture enables the adoption of multiple web protocols appropriated for the `Log Provider` implementation. HTTP[2], FTP[3] and Gopher[4] are some of the suitable candidate for this task.

Trusted Miner In our solution, `Trusted Execution Environments` are the key technologies that shelter external event logs inside an `Organization`'s system by preserving data confidentiality and integrity. The `Trusted Miner` component

[2] https://httpwg.org/specs/rfc9110.html. Accessed: July 25, 2023.
[3] https://datatracker.ietf.org/doc/html/rfc959. Accessed: July 25, 2023.
[4] https://www.rfc-editor.org/rfc/rfc1436.html. Accessed: July 25, 2023.

is a trusted application running inside the Trusted Execution Environment that makes its source code and data tamper-proof. The main goal of this component is to allow miner Organizations to execute process mining algorithms using event logs retrieved from provider Organizations, offering fair data utilization to the latters. In section 4.1, we show an high level schematization of Trusted Miners in which we distinguish four different modules: the Trusted Memory, the Log Requester, the Log Receiver, and the Log Elaborator.

Data contained in event logs belonging to provider Organizaitions are stored by miners in the Trusted Memory. This module relies on an hardware-encrypted zone of the cpu which makes event log manipuluations impossible from outside the Trusted Execution Environment. Modules of the Trusted Miner are the only entities enabled to access data stored in the Trusted Memory. Referring to our motivating scenario, the only way for the hospital to use the event logs retrieved from the specialized clinic is via the secure procedures offered by its Trusted Miner.

As we will see in section 4.2, the Log Requester and the Log Receiver are the fundamental modules that we employee during the event log exchange. Log Requesters initalize the exchange procedure and sends authenticable data requests to the Data Provision module of log providers. The Log Receiver collects event logs sent by Log Providers and store them in the Trusted Memory. This module offers methodologies that allow Log Providers to perform the remote attestation and verify the trustworthiness of the miner. According to our running example, the hospital's Log Receiver must prove its trusted nature to the clinc and the pharmaceutical company's Log Provider before receiving event logs, showing that it is actually running in a Trusted Execution Environment.

The Log Elaborator is the core module of the Trusted Miner. It collects the logic to run process mining algorithms in the Trusted Execution Environment. It support the integration of different family of tecniques such as *process discovery* [**?**], *conformance checking* [**?**] and *performance analysis* [**?**]. When activated, the Log Elaborator interacts with the Trusted Memory in order to access external event logs stored inside the Trusted Execution Environment. These are integrated with the local event log by the Log Elaborator via the merging procedure. During the merging, the Log Elaborator enriches local traces with events belonging to logs retrieved from partner Organizations. The integrity of the process mining tecniques that we implemented in the Log Elaborator is guaranteed by the Trusted Execution Environment that prevents malicious code manipulations from the running operative system. Mentioning our motivating scenario, the Log Elaborator inside the Trusted Miner of the hospital merges the traces

4.2 Workflow

Initialization

Data Exchange

Data Elaboration

5 Implementation

5.1 Event Log Generation

5.2 Trusted Miner and Log Provider

6 Evaluation

6.1 Discussion

Privacy

Security

Integreatebility

6.2 Convergence Study

Settings

Results

7 Conclusion and Future Work

Limitations:

- Both producer and consumer act fairly (so we do not expect to have injected data)
- We do not manage TEE crashes
- We assume a perfect communication channel (no loss, no snap, no corrupted bits)
- Universal clock for event timestamps (cite Event log cleaning for business process analytics by Andreas Solti)

Future Work:

- Declarative models adaptation
- Output inside the TEE, interactions through trusted applications
- Real-world event log data
- Usage policies integration

References

1. W. M. van der Aalst, "Intra-and inter-organizational process mining: Discovering processes within and between organizations," in *The Practice of Enterprise Modeling: 4th IFIP WG 8.1 Working Conference, PoEM 2011 Oslo, Norway, November 2-3, 2011 Proceedings 4*, pp. 1–11, Springer, 2011.
2. G. Elkoumy, S. A. Fahrenkrog-Petersen, M. Dumas, P. Laud, A. Pankova, and M. Weidlich, "Shareprom: A tool for privacy-preserving inter-organizational process mining.," *BPM (PhD/Demos)*, vol. 2673, pp. 72–76, 2020.
3. R. Engel, W. Krathu, M. Zapletal, C. Pichler, R. J. C. Bose, W. van der Aalst, H. Werthner, and C. Huemer, "Analyzing inter-organizational business processes: process mining and business performance analysis using electronic data interchange messages," *Information Systems and e-Business Management*, vol. 14, pp. 577–612, 2016.
4. G. Elkoumy, S. A. Fahrenkrog-Petersen, M. Dumas, P. Laud, A. Pankova, and M. Weidlich, "Secure multi-party computation for inter-organizational process mining," in *Enterprise, Business-Process and Information Systems Modeling: 21st International Conference, BPMDS 2020, 25th International Conference, EMMSAD 2020, Held at CAiSE 2020, Grenoble, France, June 8–9, 2020, Proceedings 21*, pp. 166–181, Springer, 2020.
5. J. D. Hernandez-Resendiz, E. Tello-Leal, H. M. Marin-Castro, U. M. Ramirez-Alcocer, and J. A. Mata-Torres, "Merging event logs for inter-organizational process mining," in *New Perspectives on Enterprise Decision-Making Applying Artificial Intelligence Techniques*, pp. 3–26, Springer, 2021.
6. J. Claes and G. Poels, "Merging event logs for process mining: A rule based merging method and rule suggestion algorithm," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7291–7306, 2014.
7. R. Engel and R. J. C. Bose, "A case study on analyzing inter-organizational business processes from edi messages using physical activity mining," in *2014 47th Hawaii International Conference on System Sciences*, pp. 3858–3867, IEEE, 2014.
8. R. Engel, W. Krathu, M. Zapletal, C. Pichler, W. M. van der Aalst, and H. Werthner, "Process mining for electronic data interchange," in *E-Commerce and Web Technologies: 12th International Conference, EC-Web 2011, Toulouse, France, August 30-September 1, 2011. Proceedings 12*, pp. 77–88, Springer, 2011.
9. G. Arfaoui, S. Gharout, and J. Traoré, "Trusted execution environments: A look under the hood," in *2014 2nd IEEE international conference on mobile cloud computing, services, and Engineering*, pp. 259–266, IEEE, 2014.
10. H. Xie, J. Zheng, T. He, S. Wei, and C. Hu, "Tebds: A trusted execution environment-and-blockchain-supported iot data sharing system," *Future Generation Computer Systems*, vol. 140, pp. 321–330, 2023.
11. D. Basile, C. D. Ciccio, V. Goretti, and S. Kirrane, "Blockchain based resource governance for decentralized web environments," *Frontiers in Blockchain*, vol. 6, may 2023.
12. W. M. van der Aalst, "Federated process mining: exploiting event data across organizational boundaries," in *2021 IEEE International Conference on Smart Data Services (SMDS)*, pp. 1–7, IEEE, 2021.