Security for Next-Gen Analytics for Cross-Organisation Collaboration

Laurent Gomez[1], Francesco Capano[1] and Patrick Duverger[2]
[1]*SAP Labs France, SAP Security Research, Mougins, France*
[2]*Chief Information Officer, City of Antibes, France*

Keywords: Privacy Enhancing Technique (PET), Blockchain, Trusted Execution Environment (TEE), Cross- Organisation Collaboration.

Abstract: With the global economic and energy crisis, businesses are under pressure to create more financially sustainable and environmentally-aware industries. To that extent, organizations rely on advanced analytics to optimize their business operations and mitigate risks. However, the increasing complexity of cross-organizational collaboration and ever-stricter data protection obligations pose two conflicting objectives: achieving transparency in collaborative processes - mandatory for data and process mining - while adhering to data protection obligations. In this paper, we elaborate on an approach for privacy-preserving analytics, on data shared along cross-organization collaborations. Our strategy is two-fold: (1) transparency and traceability in cross-organization collaboration, leveraging distributed ledger technologies, and (2) privacy-preserving data and process analytics, using hardware-assisted PET, Privacy Enhancing Technology. In a co-innovation with the city of Antibes, we evaluated the feasibility and performance of our approach on a public procurement use case, demonstrating a 5% decrease in late payment penalties.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

1.1 Context

As a result of the global economic and energy crisis, companies are confronted with new business imperatives for a more financially viable and sustainable industry. However, as cross-organization collaborations become more complex, organizations face increased financial burdens due to the lack of transparency.

Cross-organization collaboration can be defined as the business networks of a collaborative organization aiming at a common business objective. Legally framed by contracts, those collaborations leverage in strengths of each organization, sharing information, resources, or knowledge.

The lack of transparency and traceability slows internal transactions, impedes cross-organization process mining, or exacerbates conflict occurrence. All are caused by non-compliance with legal and contractual obligations disrupting the objective of cross-organization collaboration.

1.2 Problem Statement

The digitalization of contract management enables increased transparency and traceability. However, it comes at the expense of publicizing sensitive business information like transaction volume or contractually agreed-upon prices. This openness requirement forces organizations to share private information with various internal and external collaborators, including contractual, master, and transactional data.

Consequently, the openness imperative directly conflicts with the compliancy duty with laws, regulations, or contractual obligations. Next-gen contract management must find a way to reconcile openness and non-disclosure of sensitive business information and processes to reduce costs and mitigate inherent risks associated with those collaborations.

1.3 Approach

In this paper, we target privacy-preserving analytics on sensitive data shared along cross-organization collaborations. Our strategy is twofold:

- We implement **immutable end-to-end collaboration management** for transparency and traceability of cross-organization processes, us-

Gomez, L., Capano, F. and Duverger, P.
Security for Next-Gen Analytics for Cross-Organisation Collaboration.
DOI: 10.5220/0012044900003555
In *Proceedings of the 20th International Conference on Security and Cryptography (SECRYPT 2023)*, pages 604-609
ISBN: 978-989-758-666-8; ISSN: 2184-7711
Copyright © 2023 by SCITEPRESS – Science and Technology Publications, Lda. Under CC license (CC BY-NC-ND 4.0)

ing distributed ledger technologies, namely public permissioned distributed ledger (e.g., Blockchain (Nakamoto, 2008)).

- We enable **Privacy-preserving data and process analytics** on sensitive shared data making use of hardware-assisted PETs, Privacy Enhancing Technology, (e.g., Trusted Execution Environment (Sabt et al., 2015)).

1.3.1 Immutable End-to-End Collaboration Management

At the execution of a cross-organization collaboration, each involved stakeholder - organization_i - commits transactions and their associated attributes within a transaction distributed ledger. Each transaction is encoded as an immutable block, cryptographically linked to the previous one. This guarantees the integrity and immutability of transactions, regardless of the distributed ledger technology used. The ledger serves as a shared and decentralized database of transactions for further data and process analytics.

1.3.2 Privacy-Preserving Data and Process Analytics

Before committing a transaction and associated attributes in the ledger, organization_i encrypts it with its own managed cryptographic material. Leveraging Privacy Enhancing Technologies (PETs), we target here data and process analytics over protected data.

The remainder of this paper is organized as follows. We study state-of-the-art in section 2, followed by a detailed description of our approach in section 3. In section 4, we introduce a public procurement use case. In section 5, we evaluate the results of our approach. Finally, we summarize our findings and discuss future research directions in section 6.

2 STATE OF THE ART

In this article, we propose a novel approach for enabling privacy-preserving data and process analytics in cross-organization collaborations. We prioritize here data protection without compromising the system's functionalities. To the best of our knowledge, no similar approach have been reported in the literature.

Our approach is closely associated with Privacy Enhancing Technology (PET) (Van Blarkom et al., 2003) research field. In this section, we study state-of-the-art on PETs, categorizing them into two distincts groups: techniques relying on cryptographic primitives (crypto-based PETs) and those leveraging on hardware security mechanisms (hardware-assisted PETs).

2.1 Crypto-Based PET

Crypto-based PETs such as Homomorphic Encryption (Viand et al., 2021), Multi Party Computation (Evans et al., 2018), and Functional Encryption (Boneh et al., 2011) can be formally proven to offer privacy guarantees. However, they introduce significant resource consumption overhead and lack scalability when considering the diversity of processing capabilities required by data and process analytics.

2.2 Hardware-Assisted PET

Hardware-assisted PETs, such as Trusted Execution Environments (TEEs) (Intel, 2021), leverage hardware-based security mechanisms to guarantee private computations. However, existing implementations have drawbacks, such as reliance on a Trusted Third Party for key management and function access control, and vulnerability to side-channel attacks (Kim, 2019) (Fei et al., 2021). Solutions that provide privacy-preserving function evaluation leveraging on TEE, such as IRON (Fisch et al., 2017) and STEEL (Bhatotia et al., 2021), do not meet our decentralized security requirements as they rely on a Trusted Third Party for key management and analytic access control.

Our proposed approach allows each party to maintain ownership of its cryptographic material and analytics access control policy in cross-organization collaborations.

3 PRIVACY-PRESERVING FOR NEXT-GEN ANALYTICS

Depicted in Figure 1, we organize our overall process in three phases: data and process analytics profile management, cryptographic profile management, and data and process analytics evaluation.

3.1 Data and Process Analytic Profile Management

The *Data&Process Analytics Distributed Ledger* serves as a distributed and immutable repository for all agreed Data and Process Analytic Profile.

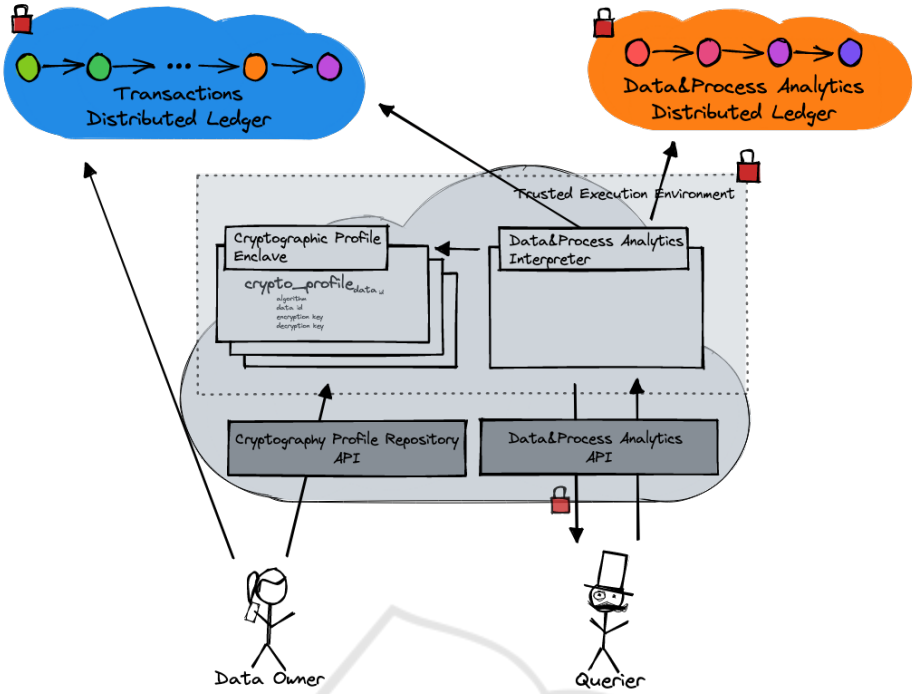

Figure 1: Architecture.

This phase includes generating and deploying the Data and Process analytic profile in the *Data and Process Analytics Distributed Ledger*. It defines each analytics' required input, pseudo-code and access control policy. Each stakeholder, owning processed data, needs to agree to deploy this profile on the *Data and Process Analytics Distributed Ledger*. We consider agreement on available analytics as out of the scope of this work. But few technologies, such as Quorum blockchain (Consensys,), can be used to deploy a profile as an immutable block based on a majority voting consensus mechanism.

3.2 Cryptographic Profile Management

All cryptographic profile are embedded within *Cryptographic Profile Enclaves*, and deployed within a joint Trusted Execution Environment.

The *Cryptographic Profile Repository API* enables the data owner to securely deploy those enclaves. This phase includes generation and deployment of cryptographic profiles within secure enclaves on a joint Trusted Execution Environment (TEE) platform.

This profile defines a cryptographic context per data type (e.g., purchase order amount) or per data group (e.g., purchase order-related information). Each stakeholder is responsible for generating and deploying those cryptographic profiles. In addition, each stakeholder must protect their encryption material (e.g., encryption key, salt). We leverage remote attestation (SSLab Georgia Tech, 2023), as defined by TEEs, for the deployment of cryptographic profiles through the *Cryptographic Profile Repository API* over a secure channel. We consider the cloud-based TEE platform to be trusted by all collaborative participants. This platform enables the deployment of cryptographic profiles within secure enclaves.

Prior to transaction commitment into the *Transactions Distributed Ledger* -transactional, master, or contractual – data are encrypted with their mapped cryptographic profile.

3.3 Data & Process Analytic Evaluation

Data owner persists any transactions, and associated -encrypted- data, to the *Transaction Distributed Ledger*. Through the *Data&Process Analytics API*, a querier requests the evaluation of an agreed Data and Process Analytics - stored in the *Data&Process Analytics Distributed Ledger*. The required data are extracted from the *Transactions Distributed Ledger*. Over local attestation, the *Data&Process Analytics Interpreter* retrieves required decryption keys from *Cryptographic Profile Enclaves*.

We evaluate Data and Process Analytics over shared encrypted data. A querier initiates a Data and Process analytic evaluation via the *Data & Process*

Analytics API. Optionally, the query provides its public key for encrypting the evaluation output. The *Data & Process Analytics API* first retrieves the Data and Process Analytic profile from the *Data&Process Analytics Distributed Ledger*. The secure enclave *Data & Process Analytics Interpreter* starts with access control enforcement based on querier authentication. The required input is collected from the transaction distributed ledger if access is granted. Over local attestation, cryptographic profiles are collected of each required data type. Required -encrypted- data is decrypted within the secure enclave, and pseudo-code is interpreted. The outcome of this evaluation is optionally encrypted with the querier public key and sent back to the querier. We consider the above process of Data and Process analytic profile evaluation by the *Data & Process Analytics Interpreter* as a contribution to this work.

4 PUBLIC PROCUREMENT USE CASE

We illustrate our approach with a public procurement use case In this cross-organization collaboration, three organizations are involved: a Public Actor, a Supplier, and a Transport Management. Public Actor requests supplies via a Purchase Order sent to a Supplier. The latter prepares requested supplies and delegates delivery to Transport Management. After validation, Public Actor pays the Supplier.

Contracts define the business interactions between public procurement stakeholders. Contractual terms and conditions, such as delivery, payment obligations (e.g., due-time delay, fixed penalties), or agreed-upon prices, are typically documented by public actors and suppliers in contracts. If any party does not fulfill its obligation accordingly, penalties are applied.

4.1 Transparency

Transparency-wise, public actors need to demonstrate efficient and optimized public funds spent while maximizing their services to citizens. A few examples of analytics requirements are as follows:

- **Obligation management** with late delivery or payment penalties triggering, computation of due penalties amount;
- **Data mining** with the evaluation of late delivery percentage, the impact of late payment penalties on global spent budget;
- **Process mining** with evaluation of average processing time per transaction;

4.2 Privacy-Preserving

Privacy-preserving wise, involved stakeholders have the non-transferable obligation to be compliant with the data protection regulation (e.g., GDPR (EU Commission, 2016), HIPAA (US Congress, 1996)) and contractual obligations. They must guarantee the protection of contractual, master, and transactional data. A few examples of sensitive shared data are as follows:

- **Contractual data**, including any trade secret, contractually agreed, such as penalties formula, due-time delays for delivery or payment, negotiated supplies prices;
- **Transactional data** such as Purchase Order details, transaction volume;
- **Master data** such as Purchase Order unitary prices or delivery details (e.g., address, contact name, email, or phone number).

5 EVALUATION

In this section, we evaluate the feasibility and performance of our approach applied to the public procurement scenario introduced in section 4.

5.1 Technical Background

We utilized Azure Cloud Confidential Computing (Microsoft, 2023) and Open Enclave SDK (OpenEnclave, 2023) to develop our demonstrator on a cloud-based TEE, reducing the attack surface of our enclaves. The *Cryptographic Profile Repository API* and *Data&Process Analytics API* were developed as REST API using Python. The distributed ledgers, *Data&Process Analytics Distributed Ledger* and *Transaction Distributed Ledger*, were deployed on a custom deployment of the public permissioned blockchain, Hyperledger (The Linux Foundation, 2023). All data was encrypted with AES-256 prior to deployment on the ledger, and Cryptographic Profile Enclaves were deployed on the Azure Cloud Computing platform. We implemented a key exchange and secure channel for exchanging Cryptographic profiles between enclaves using the local and remote attestation protocols from OpenEnclave (OpenEnclave, 2022).

5.2 Evaluation

In collaboration with city of Antibes, we evaluated our approach's impact on communication and pro-

cessing time introduced by the use of distributed ledgers and TEE enclaves.

We used approximately 8000 historical transactions for the evaluation of 5 analytics. Table 1 lists the outcome of those evaluated analytics.

Before our experimentation, the city encrypted, using AES-256, and persisted 8000 public procurement transactions within the *Transactions Distributed Ledger*. The above-mentioned analytics has been evaluated 100 times with four transaction dataset sizes: 1, 100, 1000, and 8000.

5.2.1 Communication Time Overhead

Our approach adds two potential communication time overheads that impact response time: *Data&Process Analytics API* to *Data&Process Analytics Interpreter* enclave and *Data&Process Analytics Interpreter* enclave to *Transactions Distributed Ledger*. Figure 2 shows communication time and enclave communication overheads. Enclave communication cost is negligible (a few milliseconds) due to the establishment of TIPC(TIPC Working group, 2015) sockets. However, communication between the enclave and the distributed ledger has a significant impact on performance. This occurs when the *Interpreter* gathers data from the transactions ledger for analytic evaluations. Figure 2 shows that communication overhead increases depending on the size of the collected data, up to 0.2-0.3 seconds.

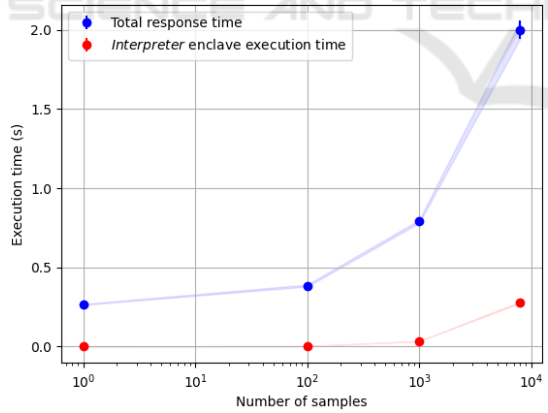

Figure 2: Communication time overhead.

5.2.2 Processing Overhead

In addition to network communication overhead, we also measured processing overhead within the interpreter enclave. Figure 3 shows the execution times for analytics on encrypted data, both inside and outside of the enclave. The gap between the two approaches grows linearly, with a larger gap after evaluating around 10^3 samples. This is due to processing overhead in the enclave, but we consider it negligible as the impact is only a few milliseconds for two years of processed transactions.

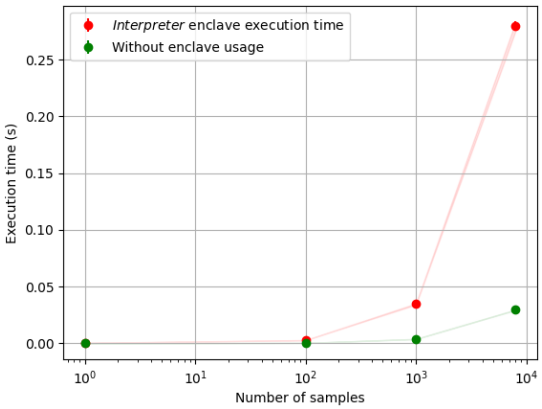

Figure 3: Processing Overhead.

5.2.3 Accuracy Loss

Hardware-assisted PET, such as Intel SGX, can improve the accuracy of privacy-preserving analytics by allowing computations on clear text data within secure enclaves. This eliminates approximation errors associated with cryptographic techniques, resulting in more accurate and sophisticated analytics while preserving privacy. Decryption within secure enclaves prevents accuracy loss, and the absence of added noise ensures the processed data and its outcome remain unaltered.

Process mining over encrypted historical data enabled the city of Antibes to detect late delivery and payment obligations triggering penalties, despite the absence of such data in the encrypted shared historical records. The *Data&Process Analytics Interpreter* used date comparison and penalty formula evaluation on encrypted data, expanding the range of processing beyond simple statistics. This capability allowed the city to perform a root cause analysis on late payments, resulting in a 5% decrease in such transactions.

6 CONCLUSION

In this paper, we elaborate a secure approach for the next-gen of analytics for cross-organisation collaboration. We target privacy-preserving data and process mining on sensitive shared data along cross-organizational transactions, improving their transparency and traceability. To that extent, we combine hardware-assisted PETs and distributed ledger technologies. With our public procurement prototype, we demonstrate the negligible impact of TEE on the eval-

Table 1: Public Procurement Evaluated Analytics.

Analytic	Description	Result
Late deliveries (%)	Percentage of over-due time deliveries	32.80 %
Late payments (%)	Percentage of over-due time payments	17.47 %
Small purchase order (%)	Percentage of Purchase Orders below 100€	65.75 %
Late deliveries impact (%)	Impact of late deliveries penalties on suppliers' total budget spent	0.36 %
Late payments impact (%)	Impact of late payments penalties on city's total budget spent	19.45%

uation of analytics over encrypted data. However, we significantly enhance the adherence to regulations and contractual security obligations. Our prototype enables city of Antibes to decrease by 5% their late payment penalties.

This work paves the way to further ML-based processing for prescriptive and predictive analytics. We foresee the deployment of analytics capabilities for both recommendations on the best course of actions at run-time and predictions on the future of evolution of those cross-organization collaborations.

ACKNOWLEDGEMENTS

This research has been conducted in the scope of the PROPOLIS(Propolis, 2023) project, jointly funded by ANR and BMBF.

REFERENCES

Bhatotia, P., Kohlweiss, M., Martinico, L., and Tselekounis, Y. (2021). Steel: composable hardware-based stateful and randomised functional encryption. In *IACR International Conference on Public-Key Cryptography*, pages 709–736. Springer.

Boneh, D., Sahai, A., and Waters, B. (2011). Functional encryption: Definitions and challenges. *Theory of Cryptography Conference*, pages 253–273.

Consensys. Quorum. https://consensys.net/quorum/.

EU Commission (2016). General Data Protection Regulation (GDPR). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679.

Evans, D., Kolesnikov, V., Rosulek, M., et al. (2018). A pragmatic introduction to secure multi-party computation. *Foundations and Trends® in Privacy and Security*, 2(2-3):70–246.

Fei, S., Yan, Z., Ding, W., and Xie, H. (2021). Security vulnerabilities of sgx and countermeasures: A survey. *ACM Computing Surveys (CSUR)*.

Fisch, B., Vinayagamurthy, D., Boneh, D., and Gorbunov, S. (2017). Iron: functional encryption using intel sgx. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 765–782.

Intel (2021). Intel Software Guard Extensions. https://www.intel.com/content/www/ architecture-and-technology/software-guard-extensions.html.

Kim, T. (2019). Sgx security. https://sgx101.gitbook.io/sgx101/sgx-security.

Microsoft (2023). Confidential Computing - Azure Solutions. https://azure.microsoft.com/en-us/solutions/confidential-compute/#overview.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Bitcoin.org*.

OpenEnclave (2022). Attestation sample OpenEnclave SDK. https://github.com/openenclave/openenclave/blob/ master/samples/attestation/README.md.

OpenEnclave (2023). OpenEnclave SDK. https://github.com/openenclave/openenclave.

Propolis (2023). PROPOLIS. https://propolis-project.eu/.

Sabt, M., Achemlal, M., and Bouabdallah, A. (2015). Trusted execution environment: What it is, and what it is not. In *2015 IEEE Trustcom/BigDataSE/ISPA*, volume 1, pages 57–64.

SSLab Georgia Tech (2023). Intel Attestation Protocol. https://sgx101.gitbook.io/sgx101/sgx-bootstrap/attestation.

The Linux Foundation (2023). Hyperledger. https://www.hyperledger.org/.

TIPC Working group (2015). TIPC Protocol Specification. http://www.tipc.io/protocol.html.

US Congress (1996). Health Insurance Portability and Accountability Act (HIPAA). https://www.hhs.gov/hipaa/for-professionals/special-topics/hipaa-administrative-simplification-regulations.

Van Blarkom, G., Borking, J. J., and Olk, J. E. (2003). Handbook of privacy and privacy-enhancing technologies. *Privacy Incorporated Software Agent (PISA) Consortium, The Hague*, 198:14.

Viand, A., Jattke, P., and Hithnawi, A. (2021). Sok: Fully homomorphic encryption compilers. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1092–1108. IEEE.