A Privacy-Preserving Data Aggregation Scheme for Fog/Cloud-Enhanced IoT Applications Using a Trusted Execution Environment

Newton Carlos Will
Computer Science Department
Federal University of Technology - Paraná
Dois Vizinhos, Brazil
will@utfpr.edu.br

***Abstract*—The use of IoT devices is increasingly present in our daily lives, as they offer many possibilities for developers and the industry to develop applications, taking advantage of their connectivity capabilities, low cost, and, often, small size. As the use of these applications is continuously increasing, the concerns about the privacy and confidentiality of the data generated by these devices also increase, since many applications share the collected data with fog and cloud servers, due to the computational constraints of the edge devices. Fog and cloud environments are used to aggregate and analyze data collected by multiple devices, allowing to summarize these data and to offer personalized services to the users. As IoT devices can collect sensitive data from users, such as personal and behavioral information, it is crucial to handle such data ensuring the privacy of their owners. Privacy-preserving data aggregation schemes are proposed in the literature, but many of them are limited to specific functions and homogeneous data or to specific contexts, such as smart metering and e-health, and there is no publicly available tool to handle heterogeneous data. This paper describes ongoing research that aims to build a generic data aggregation scheme, taking advantage of Trusted Execution Environments (TEE) to ensure data and user privacy and allowing to process heterogeneous data and perform complex computations, including the use of machine learning algorithms. We describe the system architecture, our preliminary findings, and the next steps to implement and validate our proposal.**

***Index Terms*—SGX, security, confidentiality, data processing, sensitive data, user privacy.**

I. Introduction

The Internet of Things (IoT) concept has gained much attention in the last years, due to the capacity to create various new applications by interconnecting many devices, enabling to explore new domains, such as logistics, healthcare, surveillance, industrial control, and many others. Despite the fact that IoT devices have many advantages, including cost, flexibility, and remote management, the capabilities of these devices often do not allow to perform more complex tasks due to processing, memory, and battery constraints. To overcome these limitations, integration with cloud environments can be performed [1].

One of the key features of the cloud computing model is the concept of *resource pooling*, in which workloads associated with multiple users (known as *tenants*) are typically placed in the same pool of physical resources. Cloud computing provides elastic scalability features, where the number of resources can be increased or decreased based on user demand, with minimal management effort or service provider interaction [2].

Alternatives to cloud computing are edge and fog computing. In the edge computing scenario, there is a local data processing, with the gateway located together with the IoT devices [3]. When more processing power is required, with response time constraints, local servers can be used, performing data process and storage. This scenario is defined as fog computing, which can reduce latency and decrease costs with cloud processing and communication [4]. Data processed in the edge and fog can also be sent to the cloud, in order to be analyzed and shared with other systems. Fig. 1 describes the interaction among devices, edge, fog, and cloud environments.

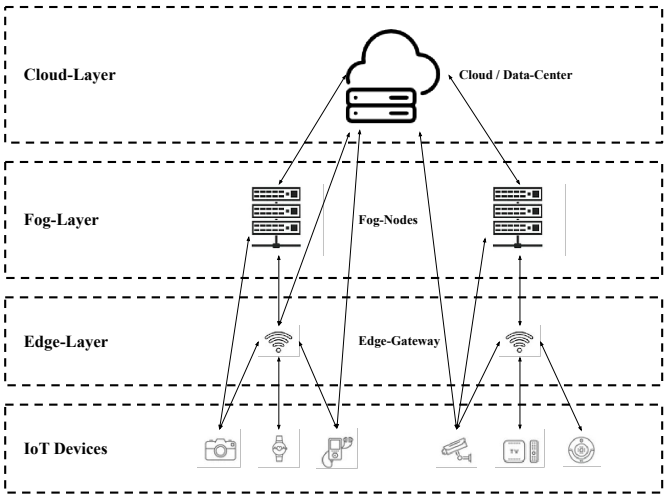

Fig. 1. Common scenarios for IoT applications. The arrows represent the possible communication between the devices and the edge, fog, and cloud layers.

Fog and cloud environments provide more computational power to summarize and analyze the data generated by the IoT devices. This integration brings many advantages to IoT

applications, but some issues concerning data privacy and confidentiality must be considered, since many IoT applications deal with sensitive data, such as medical, financial, and personal information, which can be attractive and valuable for cybercriminals.

In this sense, protecting data and user privacy is one of the most critical concerns presently, and data produced by IoT devices must be handled safely since they can reveal a lot of information about their owners. A way to deal with sensitive data in fog and cloud environments is the use of privacy-preserving data aggregation schemes, but most of them only support one type of aggregation for homogeneous IoT devices and do not allow aggregating data from heterogeneous devices [5]. Furthermore, security is an essential issue for the data aggregation process [6].

Our work seeks to fill a gap in the literature since aggregation techniques with privacy-preserving for heterogeneous data and devices are not addressed in other solutions. Many IoT systems contain heterogeneous devices operating in large, dynamic, and self-organizing networks, defined as *swarm* [7], and there is no public tool that implements and evaluates privacy-preserving mechanisms for heterogeneous data types [8].

In this paper, we propose a new scheme to perform data aggregation from heterogeneous IoT devices in fog and cloud environments with privacy-preserving constraints, by using a Trusted Execution Environment, namely Intel SGX. Such technology provides mechanisms to compute and store sensitive data in a secure way, ensuring confidentiality and integrity, and has been largely used in IoT fog- and cloud-enhanced applications [9].

The remainder of this paper is structured as follows: Section II describes the concepts of the Intel SGX technology; Section III presents the issues regarding privacy-preserving data aggregation; Section IV discuss the related work; Section V presents the system architecture of the solution proposed in this work; Section VI describes the implementation of a proof of concept; and Section VII concludes the paper and presents the next steps in our research.

II. Intel SGX

Software Guard Extensions (SGX) comprises a set of new instructions added to Intel processors that enables the developer to put sensitive data and code in a secure region, called *enclave*. Enclaves are placed in an encrypted region of memory, ensuring confidentiality and integrity. This memory has also access control primitives that prevent malicious software from accessing data, even the privileged ones, such as the operating system, hypervisor, and BIOS [10].

The main goal of the SGX is to reduce the Trusted Computing Base (TCB) to a piece of hardware and software, as shown in Fig. 2. Thus, the surface attack is reduced, resulting in software that is less likely to be compromised.

When loading the enclaves to memory, SGX is capable to detect if any part of them has been tampered with, preventing an enclave with unauthorized modifications from being loaded.

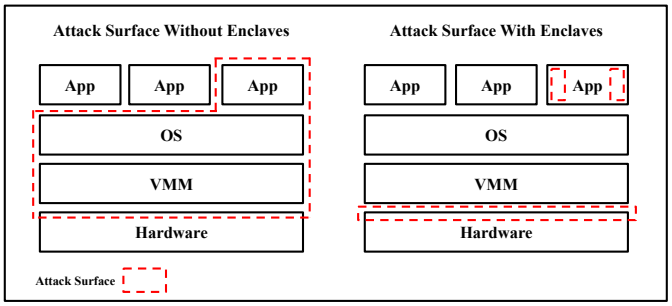

Fig. 2. Attack surface of a security-sensitive application without and with SGX enclaves [11].

Also, the enclave is linked to the application that created it, which is open to any inspection and analysis [10].

Enclaves are able to share data with each other by using an attestation mechanism, which allows an enclave to prove that it is legitimate, has not been tampered with, and was loaded correctly, allowing the creation of a secure channel for communication. Local attestation is used when both enclaves are running in the same platform, defining a symmetric key by a Diffie-Hellman key agreement procedure, authenticated by the hardware. Remote attestation is also provided, allowing to create secure communication channels over the network with third-party devices. Establishing a secure channel must be linked to the attestation process to prevent an entity from providing confidential data to an entity other than the enclave that was attested [12].

SGX also provides mechanisms to store sensitive data in persistent memory in a secure way, through the sealing process. Each enclave can request the CPU a unique key, called *sealing key*, derived from the enclave identity and the CPU itself. The sealing feature ensures that only the same enclave that seals the data can unseal them, and only using the same CPU [12].

III. Privacy-Preserving Data Aggregation

Data aggregation is the process of gathering data and presenting it in a summarized format. Data can be collected from multiple data sources with the intention of combining these data sources into a summary for data analysis. Data aggregation can provide comprehensive and accurate data to information systems, increasing the reliability and accuracy of the system and, therefore, it can be considered an important area of research.

In IoT applications, datasets collected by various devices are characterized by often containing redundant and highly correlated data. Thus, data aggregation becomes an effective mechanism to combine such redundant data and generate information, being a common operation in IoT systems.

In addition, since many systems collect and manipulate potentially sensitive data, ensuring the privacy of such data becomes a first-order concern, especially in IoT applications aided by fog and cloud environments, where the collected

data is transmitted to other providers, possibly outsourced and shared with other companies [13, 14].

IV. Related Work

Privacy-preserving data aggregation is a research topic that has been explored by several authors, with different schemes being proposed in the literature. Lu et al. [5] highlight the fact that several proposed schemes do not deal with heterogeneous data or hybrid devices, not being applicable to many IoT applications. The authors present a solution to fill this gap by using Paillier homomorphic encryption to aggregate the data and one-way hash chain techniques to perform authentication. Paillier homomorphic encryption is also used by [15], combined with signature techniques to build a scheme that is suitable in the edge layer.

Pseudonyms are used by [14] to ensure the device anonymity, as well the concept of pseudonym certificate is used to authenticate the devices. Data aggregation is performed with Paillier homomorphic encryption, with the solution being applied to fog-enhanced IoT systems. A privacy-preserving data aggregation scheme for mobile edge computing assisted IoT applications is presented by [13], protecting the privacy and providing source authentication and integrity.

Aggregation is also performed in fog and cloud environments, with [16] presenting a protocol to aggregate smart meters data in the fog by using homomorphic encryption. Wang et al. [17] also use homomorphic encryption to perform data aggregation in the fog and send them to the cloud. Another protocol is proposed by Shen et al. [18], allowing dynamic join and exit of terminal devices.

Intel SGX is used to handle sensitive IoT data in fog and cloud environments. Gremaud et al. [19] presents a middleware to perform privacy-preserving IoT data processing at the cloud, hiding data from the cloud provider and any unauthorized party. Silva et al. [20] present an architecture for data aggregation in cloud computing applied to a smart grid scenario, considering two approaches for data security and privacy: homomorphic encryption and the use of Intel SGX. Finally, privacy-aware data dissemination is proposed by [21], with data anonymization and aggregation being performed with the use of Intel SGX, which proved more feasible than homomorphic encryption.

Despite the many advances in the area and the different techniques adopted, the proposed solutions for data aggregation in fog and cloud environments do not treat heterogeneous data or data from different types of devices, focusing on a specific application context. Thus, such approaches are not applicable to many IoT applications, which manipulate and aggregate data from different sources and which allow dynamism in the inclusion and removal of devices that make up the data collection network [5].

V. Proposed Architecture

With the advent of Trusted Execution Environments (TEE), mainly the Intel SGX technology, new possibilities to perform computation over sensitive data can be explored. In this paper, we propose a new solution to perform IoT data aggregation ensuring data and user privacy, by using the enclaved execution provided by Intel SGX. We keep the data totally opaque to the cloud provider, which may be considered as an untrusted party, enabling trusted communication channels to transfer the data from the IoT devices or edge gateway to the fog or cloud environment. An illustrative schema of the architecture can be seen in Fig. 3.

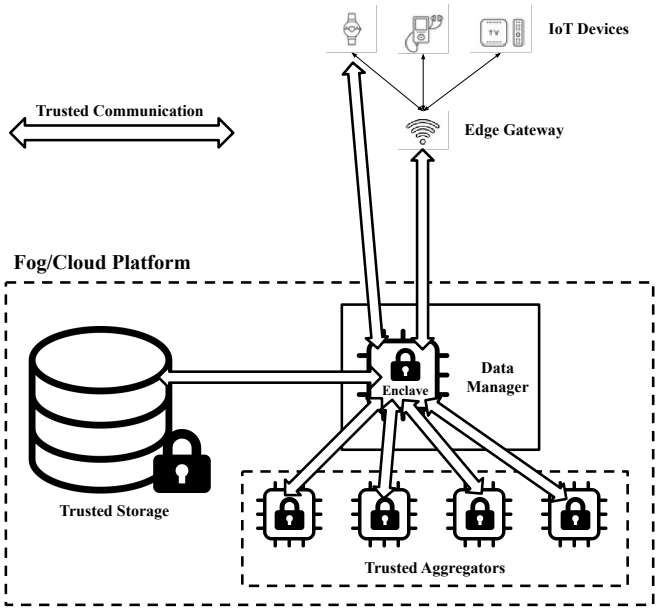

Fig. 3. The architecture of the proposed solution.

The system relies on a *data manager* to create and handle several enclaves, which we call *trusted aggregators*, that will perform the aggregation functions. The use of SGX enclaves enables to run complex functions on encrypted data at full processor speeds, ensuring data confidentiality with a reduced overhead [22]. The capabilities of Intel SGX include the execution of machine learning and neural network algorithms over private and sensitive data [23–26].

The data manager is able to handle multiple aggregators in a coordinated manner. Data manager also contains an enclave to handle sensitive information, called *enclave controller*, which will receive all data from the edge gateway or the IoT devices, choosing a specific aggregator to perform the computations. The enclave controller ensures that all requests and data received from the devices are manipulated in a trusted environment.

All the communication between the enclave controller and the edge gateway and devices is done by using a trusted channel, created by the remote attestation procedure. As described in Section II, remote attestation allows the SGX enclave to communicate securely with third parties over the network, in addition to ensuring that the enclave was correctly instantiated and was not tampered with.

Communications between the enclave controller and the trusted aggregators are also done through a secure channel,

using the local attestation procedure. This procedure ensures that all data exchanged between them are kept opaque to the untrusted provider. Also, since the data is also encrypted and handled inside the enclaves, their confidentiality is maintained, as well as the privacy of their owners.

All data received by the data manager and the results of the computations performed by the trusted aggregators can be stored in the fog/cloud environment in a secure way using the trusted storage, ensuring data confidentiality and integrity. Trusted storage relies on the sealing feature, provided by Intel SGX (see Section II), with the enclave controller requesting a sealing key to the CPU to encrypt and save all requested data in the persistent storage. Alternatively, trusted aggregators can also store data directly in fog/cloud storage.

The use of enclaved execution, provided by Intel SGX, will allow to manipulate and aggregate heterogeneous data in a feasible way, with minor overhead, when compared with other approaches described in the literature, such as homomorphic encryption [21].

The architecture described in Fig. 3 can be replicated in fog and cloud environments, allowing a multi-level data aggregation, since the cloud provider may have more computing power to perform complex data summarization and analysis, as well may be able to aggregate data coming from different fog environments.

VI. Proof of Concept

A proof of concept is currently being developed to assess the proposed architecture. Preliminary results demonstrate that, although the use of remote attestation provides strong guarantees of data confidentiality and integrity, it imposes a communication overhead between the parties that must be considered [27, 28]. SGX attestation also requires the use of the *Intel Attestation Service* (IAS) to determine if the platform the enclave is running on has SGX enabled and if the enclave is genuine. Thus, we will also evaluate the use of OPERA [29] to perform the remote attestation procedure between the enclave controller and edge gateway and IoT devices. Alternatively, data can be sent using asymmetric encryption after a key agreement, reducing the computational cost of the operation.

Enclave initialization also imposes a high computational cost, due to memory allocation, which affects the response time of the solution, since each trusted aggregator resides inside an enclave. To address this issue, we will implement enclave management techniques, such as enclave sharing [27] and pool [30], to reduce the enclave initialization overhead in the communication with the trusted aggregators.

VII. Concluding Remarks and Future Work

This paper presented a new scheme to ensure user and data privacy for data aggregation in fog and cloud environments. Our scheme relies on Trusted Execution Environment (TEE) to provide privacy and confidentiality capabilities and enables the handling of heterogeneous data generated by different devices, allowing to apply machine learning and other complex algorithms.

Although this is still a work in progress, a proof of concept of the proposed scheme is currently being developed. The implementation will be evaluated in realistic IoT scenarios, considering a strong adversarial model. A security analysis of the proposed solution would highlight its strengths and weaknesses. Also, performance tests will be carried out to measure the overhead introduced by both Intel SGX and encrypted communications.

References

[1] D. C. G. Valadares, N. C. Will, J. Caminha, A. Perkusich, Mirko Barbosa Perkusich, and K. C. Gorgonio, "Systematic literature review on the use of trusted execution environments to protect cloud/fog-based Internet of Things applications," *IEEE Access*, vol. 9, pp. 80 953–80 969, 2021.

[2] A. Rayes and S. Salam, "Fog computing," in *Internet of Things From Hype to Reality: The Road to Digitization*. Springer, 2019, pp. 155–180.

[3] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[4] C. Arivazhagan and V. Natarajan, "A survey on fog computing paradigms, challenges and opportunities in IoT," in *Proceedings of the International Conference on Communication and Signal Processing*. Chennai, India: IEEE, 2020, pp. 385–389.

[5] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT," *IEEE Access*, vol. 5, pp. 3302–3312, 2017.

[6] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of Things: A systematic review of the literature and recommendations for future research," *Journal of Network and Computer Applications*, vol. 97, pp. 23–34, 2017.

[7] B. Kuang, A. Fu, S. Yu, G. Yang, M. Su, and Y. Zhang, "ESDRA: An efficient and secure distributed remote attestation scheme for IoT swarms," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8372–8383, 2019.

[8] M. Cunha, R. Mendes, and J. P. Vilela, "A survey of privacy-preserving mechanisms for heterogeneous data types," *Computer Science Review*, vol. 41, p. 100403, 2021.

[9] N. C. Will, D. C. G. Valadares, D. F. d. S. Santos, and A. Perkusich, "Intel Software Guard Extensions in Internet of Things scenarios: A systematic mapping study," in *Proceedings of the 8th International Conference on Future Internet of Things and Cloud*. Rome, Italy: IEEE, 2021, pp. 342–349.

[10] F. McKeen, I. Alexandrovich, A. Berenzon, C. V. Rozas, H. Shafi, V. Shanbhogue, and U. R. Savagaonkar, "Innovative instructions and software model for isolated execution," in *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy*. Tel-Aviv, Israel: ACM, 2013.

[11] J. Sobchuk, S. O'Melia, D. Utin, and R. Khazan, "Leveraging Intel SGX technology to protect security-sensitive applications," in *Proceedings of the 17th International Symposium on Network Computing and Applications*. Cambridge, MA, USA: IEEE, 2018, pp. 1–5.

[12] I. Anati, S. Gueron, S. P. Johnson, and V. R. Scarlata, "Innovative technology for CPU based attestation and sealing," in *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy*. Tel-Aviv, Israel: ACM, 2013.

[13] X. Li, S. Liu, F. Wu, S. Kumari, and J. J. P. C. Rodrigues, "Privacy preserving data aggregation scheme for mobile edge computing assisted iot applications," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4755–4763, 2019.

[14] Z. Guan, Y. Zhang, L. Wu, J. Wu, J. Li, Y. Ma, and J. Hu, "APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT," *Journal of Network and Computer Applications*, vol. 125, pp. 82–92, 2019.

[15] J. Zhang, Y. Zhao, J. Wu, and B. Chen, "LVPDA: A lightweight and verifiable privacy-preserving data aggregation scheme for edge-enabled IoT," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4016–4027, 2020.

[16] F. Y. Okay and S. Ozdemir, "A secure data aggregation protocol for fog computing based smart grids," in *Proceedings of the 12th International Conference on Compatibility, Power Electronics and Power Engineering*. Doha, Qatar: IEEE, 2018, pp. 1–6.

[17] H. Wang, Z. Wang, and J. Domingo-Ferrer, "Anonymous and secure aggregation scheme in fog-based public cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 712–719, 2018.

[18] X. Shen, L. Zhu, C. Xu, K. Sharif, and R. Lu, "A privacy-preserving data aggregation scheme for dynamic groups in fog computing," *Information Sciences*, vol. 514, pp. 118–130, 2020.

[19] P. Gremaud, A. Durand, and J. Pasquier, "Privacy-preserving IoT cloud data processing using SGX," in *Proceedings of the 9th International Conference on the Internet of Things*. Bilbao, Spain: ACM, 2019.

[20] L. V. Silva, P. Barbosa, R. Marinho, and A. Brito, "Security and privacy aware data aggregation on cloud computing," *Journal of Internet Services and Applications*, vol. 9, no. 1, 2018.

[21] L. Sampaio, F. Silva, A. Souza, A. Brito, and P. Felber, "Secure and privacy-aware data dissemination for cloud-based applications," in *Proceedings of the 10th International Conference on Utility and Cloud Computing*. Austin, TX, USA: ACM, 2017, pp. 47–56.

[22] B. Fisch, D. Vinayagamurthy, D. Boneh, and S. Gorbunov, "IRON: Functional encryption using Intel SGX," in *Proceedings of the 24th Conference on Computer and Communications Security*. Dallas, TX, USA: ACM, 2017, pp. 765–782.

[23] O. Ohrimenko, F. Schuster, C. Fournet, A. Mehta, S. Nowozin, K. Vaswani, and M. Costa, "Oblivious multi-party machine learning on trusted processors," in *Proceedings of the 25th USENIX Conference on Security Symposium*. Austin, TX, USA: USENIX Association, 2016, pp. 619–636.

[24] F. Tramèr and D. Boneh, "Slalom: Fast, verifiable and private execution of neural networks in trusted hardware," in *Proceedings of the 7th International Conference on Learning Representations*. New Orleans, LA, USA: ICLR, 2019.

[25] D. L. Quoc, F. Gregor, S. Arnautov, R. Kunkel, P. Bhatotia, and C. Fetzer, "SecureTF: A secure TensorFlow framework," in *Proceedings of the 21st International Middleware Conference*. Delft, Netherlands: ACM, 2020, pp. 44–59.

[26] Y. Liang, D. O'Keeffe, and N. Sastry, "PAIGE: Towards a hybrid-edge design for privacy-preserving intelligent personal assistants," in *Proceedings of the 3rd International Workshop on Edge Systems*. Heraklion, Crete: ACM, 2020, pp. 55–60.

[27] N. C. Will and C. A. Maziero, "Using a shared SGX enclave in the UNIX PAM authentication service," in *Proceedings of the 14th Annual IEEE International Systems Conference*. Montreal, QC, Canada: IEEE, 2020, pp. 1–7.

[28] N. C. Will, T. Heinrich, A. B. Viescinski, and C. A. Maziero, "Trusted inter-process communication using hardware enclaves," in *Proceedings of the 15th Annual IEEE International Systems Conference*. Vancouver, BC, Canada: IEEE, 2021, pp. 1–7.

[29] G. Chen, Y. Zhang, and T.-H. Lai, "OPERA: Open remote attestation for Intel's secure enclaves," in *Proceedings of the 26th Conference on Computer and Communications Security*. London, UK: ACM, 2019, pp. 2317–2331.

[30] D. Li, R. Lin, L. Tang, H. Liu, and Y. Tang, "SGXPool: Improving the performance of enclave creation in the cloud," *Transactions on Emerging Telecommunications Technologies*, p. e3735, 2019.