Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Practical hybrid confidentiality-based analytics framework with Intel SGX[☆]

Abdulatif Alabdulatif

Department of Computer Science, Qassim University, Buraydah 51452, Saudi Arabia

ARTICLE INFO

Article history:
Received 4 July 2020
Received in revised form 19 June 2021
Accepted 14 July 2021
Available online 21 July 2021

Keywords:
Cloud computing
Data confidentiality
Data clustering
Intel SGX
Internet of Things

ABSTRACT

Massive cloud infrastructure capabilities, including efficient, scalable, and elastic computing resources, have led to a widespread adoption of Internet of Things (IoT) cloud-enabled services. This involves giving complete control to cloud service providers (CSPs) of sensitive IoT data by moving data storage and processing in cloud. An efficient and lightweight advanced encryption standard (AES) cryptosystem can play a major role in protecting IoT data from exposure to CSPs by protecting the privacy of outsourced data. However, AES lacks computation capabilities, which is a critical factor that prevents individuals and organizations from taking full advantage of cloud computing services. When Intel software guard extensions (SGX) is used with AES cryptosystem, the developing framework can provide a practical solution to build a confidentiality-based data analytics framework for IoT-enabled applications in various domains. In this paper, a privacy-preserving data analytics framework is developed that relies on a hybrid-integrated approach, in which both software- and hardware-based solutions are applied to ensure confidentiality and process-sensitive outsourced data in the cloud environment.

1. Introduction

The advent of Internet of Things (IoT) and edge computing has opened numerous dimensions in technology and prompted researchers to innovate at a rapid rate. IoT technology is developing quickly and has introduced serious concerns about data privacy and integrity. With IoT, the volume of data production and the sharing of data among worldwide networks is unparalleled. As more organizations, private and public, are acquiring IoT to provide solutions in health care, sustainability and other vital sectors, the need for cloud adoption is also increasing. They are bound to obtain the cloud services for storing, managing, and processing massive amounts of data. The cloud services shorten the delivery time for solutions, thereby increasing productivity. Another significant benefit is the analysis and visualization of data for timely and informed decisions, promoting efficiency.

With all these advantages of cloud ecosystem, there is an increasing number of attacks and risks associated with it that can lead to the exposition of highly sensitive data. This creates additional challenges to fundamental aspects of data confidentiality, availability, and integrity (Zissis and Lekkas, 2012). Further, immense dependence on third-party cloud providers presents a risk of corruption, illegal exposure, and misuse of organization-owned data (Sundareswaran et al., 2012; Ren et al., 2012). The extant literature confers different strategies and frameworks to eradicate the problem of data protection and preservation in an outsourced (public cloud-based) environment. The techniques include strict access-control rules, implementation of different anonymization methods and application of multi-party computation (MPC) (Atallah et al., 2001; Wang et al., 2010; Zhou et al., 2011; Chadwick and Fatema, 2012; Backes et al., 2013; Li et al., 2014). However, these techniques are limited to providing privacy-preservation solutions in a specific context, excluding the power of data computation. Even if they possess the computational capability, they are either not intelligent enough or too expensive to provide constructive data analysis for informed decisions.

The objective of this paper is to develop a practical and efficient framework for the adaption of confidentiality-based data analysis in various domains in the realm of IoT. The developed framework aims to build a hybrid privacy-preservation solution that combines both software- and hardware-based techniques to maintain data confidentiality in volatile and untrusted cloud environments. The framework comprises techniques, including advanced encryption standard (AES) (Nechvatal et al., 2001) and Intel as software guard extensions (SGX) (McKeen et al., 2013). The practical implications of AES cipher are acknowledged worldwide with regard to protection of digital data, but it does not encompass analytical computation capabilities. An alternative is homomorphic cryptosystems. However, these are either impractical or cost heavy at a large scale. The latest versions of Intel processor generations – starting from 6*th* to the currently 10*th*

[☆] Editor: Raffaela Mirandola.
E-mail address: ab.alabdulatif@qu.edu.sa.

https://doi.org/10.1016/j.jss.2021.111045
0164-1212/© 2021 Elsevier Inc. All rights reserved.

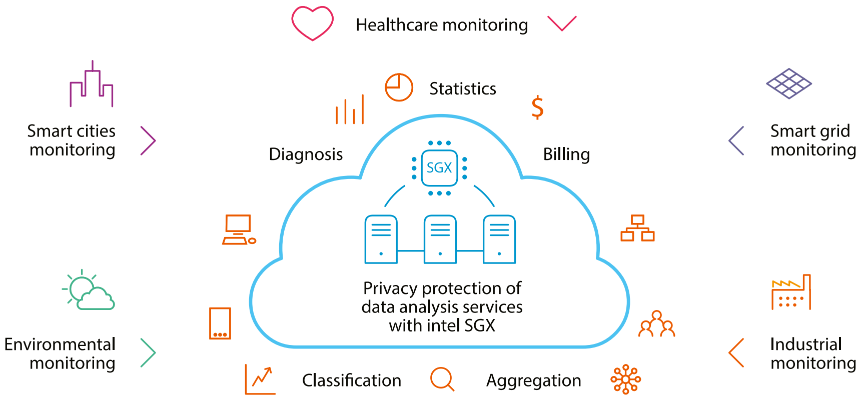


Fig. 1. Overview of a secure data analytic approach for IoT cloud-enabled framework using Intel SGX.

generation – come with the Intel SGX component that has a security feature developed to ensure the confidentiality of outsourced data at the hardware level. To overcome these limitations, SGX provides the migration of processing and data storage to an isolated memory compartment to perform computations securely without compromising data confidentiality. This embedded framework can be beneficial for end-to-end confidentiality-based data computations across IoT domains, such as health care and smart-grid applications. Fig. 1 represents a blueprint of the proposed secure data analytic framework. Applications that require processing sensitive data in various domains can benefit from the proposed framework, such as e-health diagnosis and assisted-living systems, through which patients' sensitive data can be processed efficiently while ensuring confidentiality. Further, industrial-scale applications (e.g., machine process and smart-grid monitoring systems) generate sensitive data from an industrial espionage perspective, in which disclosing this data can reveal sensitive customer data. These realistic scenarios of possible sensitive data disclosure can be eliminated when the data are stored and processed based on the proposed hybrid confidentiality-based analysis framework.

1.1. Motivation

According to Right Scale's cloud survey, Flexera (2019), 91% of enterprises outlined public cloud adoption in 2019 alone. According to Gartner report, the public cloud market investment is expected to increase by 17% in 2020 to reach 266.4 billion up from 227.8 billion in 2019. This shows the impact of rapid migration of cloud services, especially for small- and medium-sized enterprises as they equip them with essential resources for data storage and development within a small budget. While there is no doubt of the potential of cloud computing, offering cost-effective and reliable resources to organizations, several security and privacy concerns in the cloud ecosystem need to be addressed (Grobauer et al., 2010). With IoT in the frame, the need to develop privacy-preservation frameworks focused on processing and exchange of data to and from cloud resources has become of prime importance to ensure the protection of sensitive data. Ensuring the privacy of migrating data is critical to the realization of the full potential and advantages of cloud resources.

1.2. Contributions

The main contributions of this paper are as follows.

1. The development of a practical and hybrid confidentiality-based data analytics framework that combines the software AES cryptosystem and hardware Intel SGX-based security solutions to ensure end-to-end privacy protection at all phases of data communication, processing, and storage.
2. The evaluation of the developed framework in terms of analysis performance and accuracy. The experimental outcomes show that the proposed framework achieves a high level of accuracy of the overall analysis process similar to the insecure version of analysis tasks while ensuring full confidentiality protection for the data being processed in cloud computing.

The rest of the paper is further divided into the following. The literature review is presented in Section 2. The architecture of the developed framework is shown in Section 3. The threat model and applied machine-learning techniques are explained in Sections 4 and 5. Section 6 presents the security discussion, while Section 7 focuses on experimental evaluation. The concluding remarks are presented in Section 8.

2. Literature review

This section presents the prevailing research entailing secure data analytics techniques and Intel SGX implications.

Several approaches are adopted by researchers for preservation of privacy in data analytics models. The randomization- and cryptography-based approaches are widely utilized. Randomization-based approaches mask the data by adding random noise, thereby protecting data in processing phase (Agrawal and Srikant, 2000; Du and Zhan, 2003). However, to mask the data, these approaches also reduce the analytical accuracy by tampering the original data with noise Patel et al. (2015). The evidence of formal methods for security provisioning is also lacking. Conversely, the cryptography-based approaches lean on the MPC for data analysis (Goldreich, 2005). Though the discussed cryptography approaches can achieve a high level of privacy provisioning, the overhead costs and increased computation complexity are inevitable. The authors of Inan et al. (2007), Doganay et al. (2008) and Rivest et al. (1978) discussed three cryptography techniques: oblivious transfer, secret sharing, and homomorphic encryption. Oblivious transfer and secret sharing are not applicable for larger datasets because of high computation and communication costs (Duan and Canny, 2014). In contrast, homomorphic encryption techniques can perform complex computations on encrypted datasets and have two categories, as mentioned in Gamal (1985) and Gentry and Halevi (2011) (i.e., somewhat homomorphic encryption and fully homomorphic encryption). However, it is also deemed impractical at a large scale because of the increased cost and complexity. This paper focuses on developing a practical hybrid-analytical framework that will take advantage of both software- and hardware-based solutions. Advanced Encryption Standard (AES) (Daemen and Rijmen, 2020) is a well-known cryptosystem that has been proven and adopted world wide.

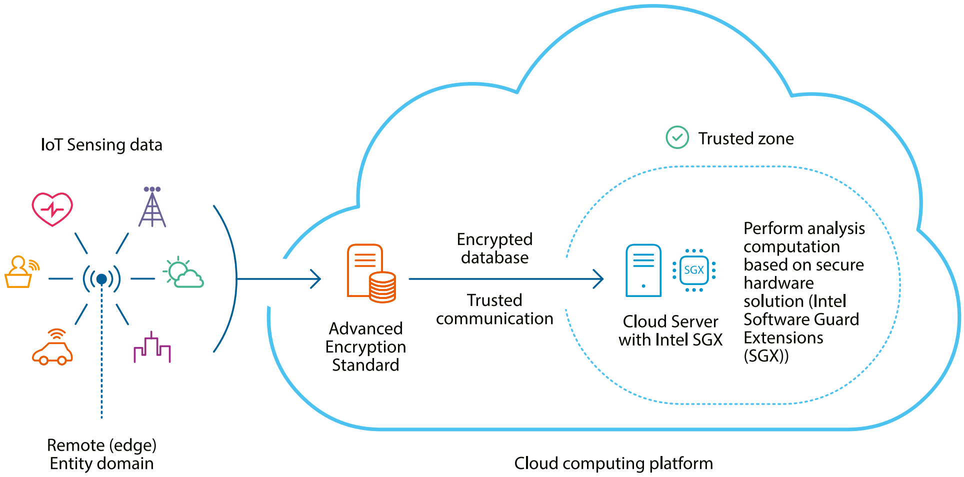


Fig. 2. Overview of the proposed hybrid confidentiality-based analytics framework for IoT cloud-enabled framework using Intel SGX.

AES cryptosystem can be used effectively to protect sensitive data, while it is at rest, or during transmission between different entities. Several approaches have been developed to enhance the efficiency of AES cryptosystems as in Oukili and Bri (2017) and Rao Rupanagudi et al. (2019), Langenberg et al. (2020). AES cryptosystems have been applied in various domains, such as healthcare and smart grids, to ensure the confidentiality of sensitive data. Recently, there has been a shift toward developing hardware-based solutions for providing protection. The aim is to add another layer at the hardware level to enhance the secrecy of data processing. These solutions are termed trusted execution environments (TEE). The Intel SGX is leveling up as a competent TEE that can provide elite privacy with reduced costs associated with data analytic computations in cloud environment. The authors of Schuster et al. (2015) explained how SGX has been applied in the Hadoop MapReduce framework for big data processing. The application of Intel SGX was also described by Zheng et al. (2017) as building a distributed data analytics service with oblivious computing. In Hunt et al. (2018), it was stated that Ryoan – a distributed sandbox specific to untrusted computations on sensitive data – has utilized SGX to improve its own effectiveness and security. As observed in previous research, there are several standalone solutions to overcome the problem of privacy-preserving analytic services. However, this paper has presented a practical hybrid approach that combines software- and hardware-based framework to provide end-to-end protection in the IoT outsourced data analytics environment. Unlike the existing solution, the developed framework aims to support the efficient implementation of various advanced analytics models, in a completely automated cloud-based platform, while taking full advantage of a cloud-computing environment, including storage and processing resources, that in turn will offer unlimited capabilities for adapting various analytical service applications, without compromising data privacy.

3. Hybrid confidentiality-based analytics framework

This section presents the proposed hybrid confidentiality-based analytics framework. This involves describing the entities, their roles, and how the entities interact to accomplish analysis tasks of sensitive IoT data in a privacy-preservation manner in the cloud.

The architecture of the proposed framework has three main entities:

- **Remote (edge) entity:** This is the data source. It can be either an end-user or a sensor-enabled IoT device in which data are collected and later disseminated to cloud storage.
- **Cloud storage entity:** This is the storage place for the data coming from edge devices. The data are in encrypted form, using an AES cryptosystem.
- **Analytic engine entity:** This is the fundamental entity of the proposed framework. In this entity, the encrypted data in cloud storage are manipulated using data-clustering techniques.

The framework entities collaborate to aggregate, store, and perform data analysis tasks while providing end-to-end privacy. The developed framework comprises two main zones of the developed framework, including a trusted zone (trusted zone as shown in Fig. 2). In the trusted zone, an isolated SGX is used to perform analysis tasks for applied analytic models including KMC and FCMC algorithms. For this, ECALL functions are used as a trusted component of SGX architecture to implement analytic models. The untrusted zone is assumed to be completely exposed to the adversary. Therefore, the AES cryptosystem (assuming the cryptosystem parameter initialization occurs in the secure remote edge entity) and the aggregated data from the remote edge entity that are transmitted for processing inside the SGX enclave. The remote (edge) entity can retrieve analysis results, for which OCALL functions are employed.

Regarding the communication channel between the trusted and untrusted zones, the remote attestation, an advanced feature of Intel SGX, plays a critical role to established an authorized communication channel between the SGX enclave and the remote (edge) entity to exchange encryption/decryption parameters and to facilitate any further data exchange, as shown in Fig. 3. The remote attestation ensures a secure communication channel for sending sensitive collected to cloud storage and retrieving analysis results. The remote attestation includes three main services: verifying the identity of the analysis services within an SGX enclave, verifying their correctness (ensuring they have not been tampered with), and ensuring that analysis services run securely within an enclave on an Intel SGX-enabled platform. After the remote attestation process is completed, the encrypted data are sent to the cloud storage entity. The analytic engine entity can complete data processing independently. Data owners (individuals or enterprises) can retrieve the encrypted result through the cloud resource and present it to the beneficiaries through dedicated and secure sites. Later in Section 5 , the data analytic entity is discussed in detail. The overall workflow model is shown in Fig. 3.

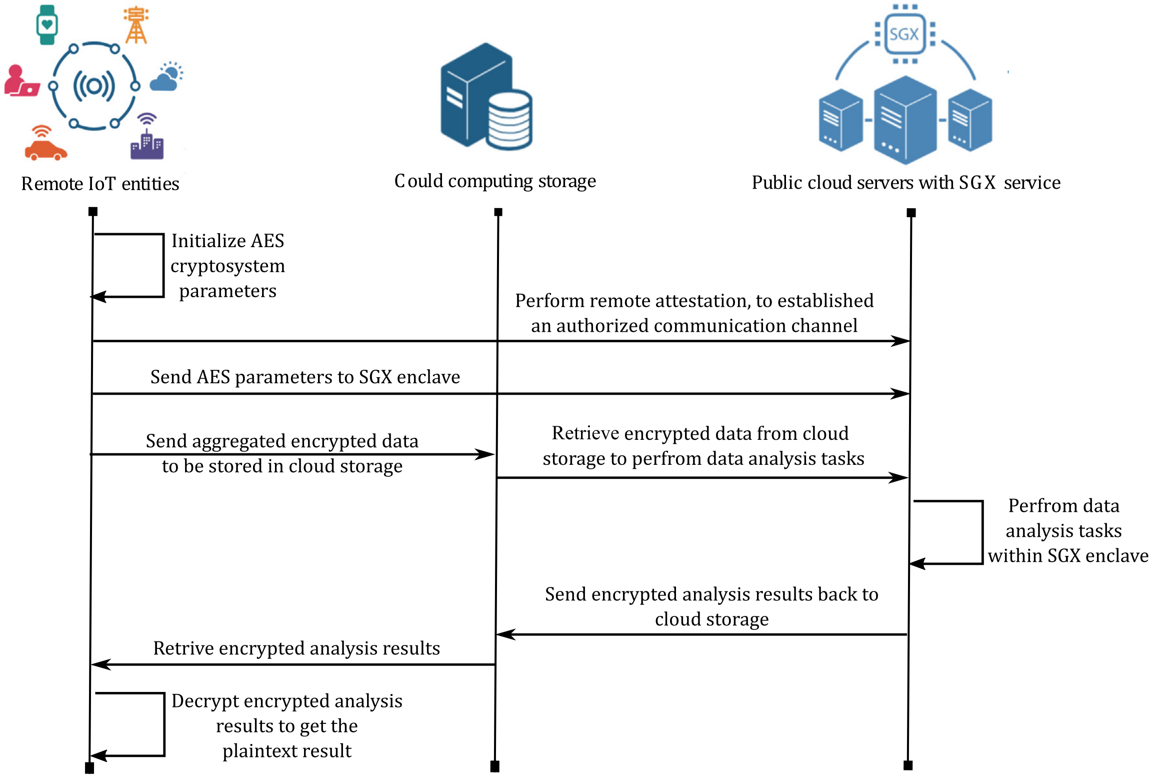

Remote IoT entities
Could computing storage
Public cloud servers with SGX service
SGX
Initialize AES cryptosystem parameters
Perform remote attestation, to established an authorized communication channel
Send AES parameters to SGX enclave
Send aggregated encrypted data to be stored in cloud storage
Retrieve encrypted data from cloud storage to perfrom data analysis tasks
Perfrom data analysis tasks within SGX enclave
Send encrypted analysis results back to cloud storage
Retrive encrypted analysis results
Decrypt encrypted analysis results to get the plaintext result

Fig. 3. Workflow model of the proposed hybrid privacy-preserving analytics framework.

4. Threat model

Before discussing the entities for the proposed framework, an assumption is made to shape the threat model—that the remote entity (i.e., end-user and edge devices) are secure to collect and receive the sensing IoT data. The rest of the model entities are vulnerable to internal and external threats. Therefore, identification of a security mechanism is essential to make the proposed framework resilient enough to withstand any compromise. This section will shed light on the way users' sensitive data and associated analytical operations will be protected through the complete lifecycle of end-to-end communication in IoT ecosystem.

4.1. Remote (edge) entity and communication channel

It has already been stated that the communication channel to and from the remote entity is not secure, despite the remote devices being secure themselves. It is essential to transfer data between the devices and storage entity in an encrypted form. To achieve this, a privacy-protection mechanism must be devised to exchange the highly sensitive information between the remote entities and Intel SGX enclave. Remote attestation can establish a secure communication channel with the remote entity. This enables the remote secure entity to transfer AES cryptographic primitives to the SGX enclave securely. It is assumed that the adversary cannot compromise the secure enclaves and their relevant keys—in this case, seal, and attestation keys. Advanced side-channel attacks, as in Chen et al. (2020) and Murdock et al. (2020), can be prevented by applying current defense techniques, as in Orenbach et al. (2020). However, this concern, along with physical and denial-of-service attacks on the remote entities, are beyond the scope of this article.

4.2. Data analytic entity

As discussed previously, the processing component of the proposed framework, the analytic engine entity, is used to perform the computational tasks. The primary feature of the proposed framework is that the computational tasks will be performed inside the Intel SGX architecture. It is also assume that the computations are processed inside the SGX enclave environment. It is further assumed that the cloud service provider (CSP) is a semi-honest party that follows framework transactions but attempts to gain more information than is allowed. The SGX enclaves hosted by CSPs are assumed to be isolated completely from BIOS, I/O, and even power of cloud servers, which are considered potentially untrustworthy. Further, an adversary may control computing resources or software, such as operating systems or hypervisors, to attack the protected analysis processes. Therefore, it is assumed that the analysis functions that run inside the enclaves are the only trusted components. The analytic based clustering computations are only dependent on built-in C/C++ libraries within SGX enclave environments. Particularly, the only computations implemented are standard arithmetic operations supplemented with exponentiation and polynomial evaluations of the initial inputs, along with intermediate results through which SGX enclaves completely assist these operations. Therefore, assuming that the SGX internal state is secure implies that the analysis computations processing inside SGX enclave are also secure.

5. Analytic services-based data clustering

Data-clustering analysis is used to categorize objects (data points) that share similar properties into different groups called clusters. For initial exploration of input data, data clustering is deemed a popular technique. It is used in various fields, including image analysis, pattern recognition, information retrieval and bioinformatics. In this paper, two principal centroid-based

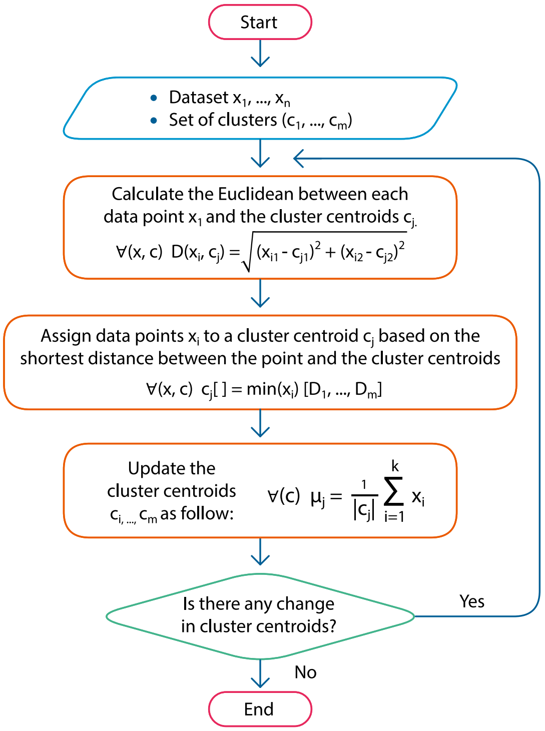

Start
Dataset x1, ..., xn
Set of clusters (c1, ..., cm)
Calculate the Euclidean between each data point xi and the cluster centroids cj.
∀(x, c) D(xi, cj) = √((xi1 - cj1)² + (xi2 - cj2)²)
Assign data points xi to a cluster centroid cj based on the shortest distance between the point and the cluster centroids
∀(x, c) cj[ ] = min(xi) [D1, ..., Dm]
Update the cluster centroids ci, ..., cm as follow:
∀(c) μj = 1/|cj| Σ_{i=1}^{k} xi
Is there any change in cluster centroids?
Yes
No
End

Fig. 4. The procedural steps of K-means clustering algorithm.

clustering algorithms are applied as proof of concept for the proposed model, including K-means clustering (KMC) and fuzzy C-means (FCM) clustering algorithms. The procedural steps for both algorithms are illustrated next.

KMC can be accomplished as follows and is diagrammatically presented in Fig. 4.

1. Let $x_1, \ldots, x_n$ be a set of two-dimensional data points. The algorithm randomly selects a set of cluster centroids $c_1, \ldots, c_m$.
2. Calculate the Euclidean distance between each data point x_i and the cluster centroids c_j.

$$\forall(x, c) \quad D(x_i, c_j) = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2} \tag{1}$$

3. Assign data points x_i to a cluster centroid c_j based on the shortest distance between the point and the cluster centroids.

$$\forall(x, c) \quad c_j[\ ] = min(x_i)[D_1, \ldots, D_m] \tag{2}$$

4. Update the cluster centroids $c_i, \ldots, c_m$.

$$\forall(c) \quad \mu_j = \frac{1}{|c_j|}\sum_{i=1}^{k} x_i \tag{3}$$

where k is the number of data points that are assigned to a cluster centroid c_j and μ_j is the updated mean of a cluster centroid c_j.
5. Repeat Steps 2,3 and 4 until there is no longer change in the updated cluster centroids (see Fig. 4).

The FCM clustering algorithm can be accomplished as follows and is diagrammatically presented in Fig. 5.

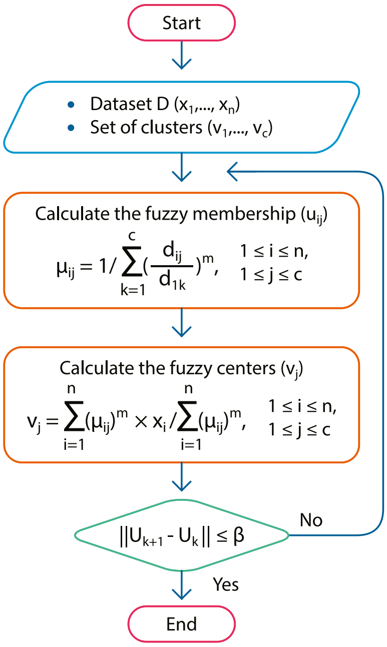

Start
Dataset D (x1,..., xn)
Set of clusters (v1,..., vc)
Calculate the fuzzy membership (uij)
μij = 1/Σ_{k=1}^{c}(dij/d1k)^m, 1 ≤ i ≤ n, 1 ≤ j ≤ c
Calculate the fuzzy centers (vj)
vj = Σ_{i=1}^{n}(μij)^m × xi / Σ_{i=1}^{n}(μij)^m, 1 ≤ i ≤ n, 1 ≤ j ≤ c
||Uk+1 - Uk|| ≤ β
No
Yes
End

Fig. 5. The procedural steps of fuzzy c-means clustering algorithm.

1. Data objects are assigned to possible clusters based on calculated membership matrices.

$$\mu_{ij} = 1/\sum_{k=1}^{c}(\frac{d_{ij}}{d_{1k}})^m \tag{4}$$

where μ_{ij} is a membership value between a data object i and a cluster centroid j. d_{ij} is an Euclidean distance between a data object i and a cluster centroid j as shown in Eq. (1).
2. Cluster centroids are updated by calculating the new means of data objects in the current clusters through the following function:

$$v_j = \sum_{i=1}^{n}(u_{ij})^m x_i / \sum_{i=1}^{n}(u_{ij})^m \tag{5}$$

where v_j is the jth cluster.
The membership values of data points and cluster centroids are updated based on Eqs. (4) and (5) until the following condition is satisfied:

$$\|U^{k+1} - U^k\| < \beta \tag{6}$$

where U is $(\mu)_{n*c}$ the fuzzy membership matrix and β is the termination criterion value that is pre-determined.

6. Security discussion

The developed hybrid privacy-preservation analysis framework aims to protect the privacy of aggregated IoT-based data and perform analysis tasks securely to prevent any malicious activities. Thus, the developed framework is secured against the threat model. In the event of an eavesdropping-based attack on the communication channel between remote entities and Intel SGX enclaves, a possible adversary could only intercept protected data through encryption, when an AES cryptosystem is applied on aggregated sensed data upon receipt to ensure its confidentiality. Further, the injection of illegitimate key material during communication can be another attack that also not possible for the attacker with Intel's SGX attestation process. The supporting defense layer effectively mitigates such vulnerabilities. This type of compromise is sometimes referred to as the Eve mechanism and was first observed as a vulnerability for naive Diffie–Hellman.

In the case of eavesdropping attacks targeting Intel SGX enclaves, the only known feasible methods to eavesdrop the sensitive data from protected the SGX enclave memory are the specter techniques, such as an adversary being able to launch side-channel attacks. Developed schemes, SCONE (Arnautov et al., 2016) and Varys (Oleksenko et al., 2018) can be deployed to overcome such attacks. Moreover, the patterns of memory access can compromise the privacy of data during data exchange and inside enclave (Sasy et al., 2018). Therefore, analytic models, such as machine-learning algorithms, can be implemented based on oblivious techniques to eliminate and execute data-dependent patterns (Ohrimenko et al., 2016). After discussing the security of individual entities in the proposed framework, the research can conclude that the entire system is secure. There is no computationally feasible mechanism to extract either data or results from the system, except with negligible probability.

7. Experimental evaluation

In this section, a set of varying experiments are conducted to assess the functionality and performance of the proposed framework. For these experiments, the primary data mining algorithms used are KMC and FCMC algorithms. The performance of adapted AES cryptosystem and communication overhead of exchanging encrypted data between IoT device (in this case, Raspberry Pi node) and Intel SGX enclave are evaluated in detail. Furthermore, clustering-based algorithms are implemented and used for plaintext and ciphertext versions comparison. The plaintext implementations are used as a baseline against the measurement of encrypted system. Two fundamental questions are asked:

1. Do the developed privacy-preservation analytic models (KMC and FCMC algorithms) achieve high level of analytic accuracy?
2. What are the relative performance overheads of the developed privacy-preservation analytic models?

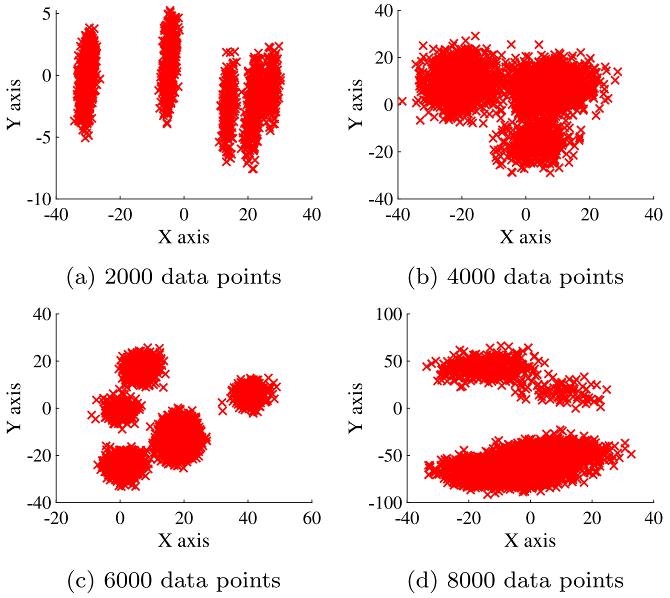
This section outlines the results obtained after series of experiments with observed comparisons between functionality and performance.

7.1. Datasets

The developed framework is evaluated using a public set of benchmark clustering datasets. These datasets are specifically designed for cluster analysis and consider varying characteristics (Franti and Virmajoki, 2006). They are represented in Fig. 6. The datasets consist of 2000, 4000, 6000, and 8000 two-dimensional data points with corresponding class labels and numerous 12 centroid clusters with different degrees of overlap. To demonstrate various aspects of the proposed framework, the datasets are divided into subsets to examine the analytic accuracy and performance overheads with varying dataset sizes.

7.2. Experimental setup

To demonstrate the experimental evaluation, a server on Microsoft Azure is deployed. the DCsv2 series machines is used, which offers SGX-enabled processors. Intel®Xeon CPU®E-2288G @ 3.70 GHz with 8 cores and 32 GiB RAM, running on Ubuntu 20.04 OS is used with a processor supports 256MB of enclave size (a total usable memory of 168MB). Moreover, Raspberry Pi 3 with 4 GB memory is used to collect and send aggregated data to the Intel SGX enclave. It is of interest to measure the performance and functionality of a complete developed encrypted-based data analytic framework and the corresponding plaintext version of analytic models.



(a) 2000 data points

(b) 4000 data points

(c) 6000 data points

(d) 8000 data points

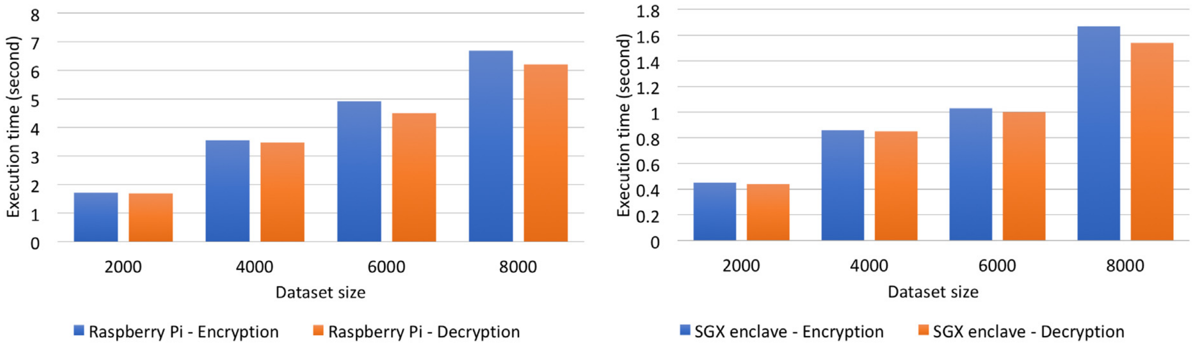
Fig. 6. The distribution of two-dimensional synthetic datasets. The datasets consist of 2000, 4000, 6000, and 8000 two-dimensional data points with corresponding class labels and varying number of cluster centroids with different degrees of cluster overlap.

The experiments comprise several phases. First, in the initialization phase, the AES cryptosystem encryption/decryption key material is generated. Second, during the key sharing phase, remote attestation is enabled to transfer key material. Third, during the encryption phase, the datasets are encrypted in the remote IoT entity. Fourth, in the transmission phase, the encrypted data are sent to the secure Intel SGX processing unit. Fifth, during the data analysis phase, the Intel SGX processing unit decrypts the data that are transferred in the second phase and performs the analysis tasks before encrypting the analysis results. Sixth, during the receiver phase, the encrypted results are transmitted back to the remote entity. In the final phase, the results are decrypted for any further processing tasks in the remote secure entity.

7.3. Performance metrics

The performance evaluation demonstrates two main criteria: analysis task accuracy and performance overheads. Fig. 7 shows the extracted execution times for the developed privacy-preservation analytic framework for both encryption and decryption processes with varying dataset sizes for both IoT-based Raspberry Pi and cloud-based SGX enclaves. Overall, IoT-based Raspberry Pi takes longer to process compared with cloud-based Intel SGX enclaves because of the limited resource capabilities of IoT-based devices. Further, it is observed that the developed privacy-preservation analytic framework and corresponding plaintext versions of KMC and FCMC algorithms produce identical analysis results regarding analytic accuracy. The result is as expected since the presence of encryption in each part of the data transmission and data receiver phases will not modify the values of the raw data. Moreover, the analysis processing of the developed framework is performed in plaintext version inside the SGX enclave, which results in similar analysis results.

From the performance perspective, the notable differences can be observed in KMC and FCM algorithms' execution time, including data encryption at remote entity, data transmission,



(a) Raspberry Pi

(b) Cloud-based SGX enclave

Fig. 7. Execution time of AES encryption and decryption processes in both IoT-based Raspberry Pi and cloud-based SGX enclaves with varying dataset sizes.

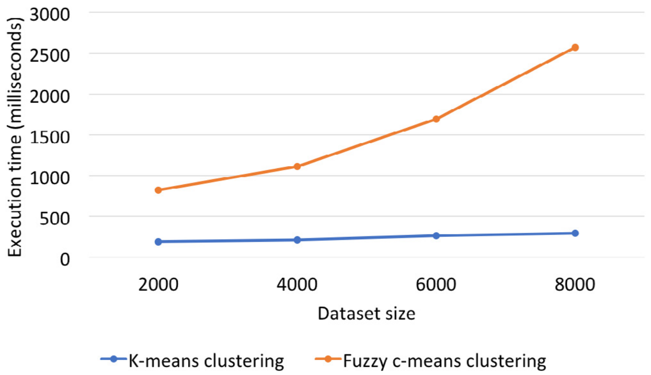


Fig. 8. Execution time for processing privacy-preservation KMC and FCMC algorithms with varying dataset sizes.

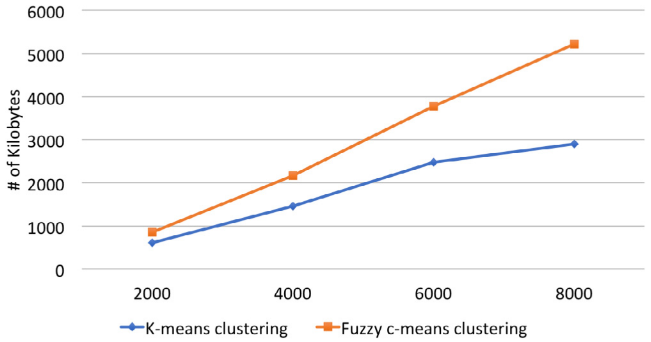


Fig. 9. The memory usage of Intel SGX enclave with varying dataset sizes.

decryption, and analysis tasks, and finally send the encrypted results back to secure remote entity. This is directly proportional to the dataset size and number of clusters. These differences are represented in Fig. 8. For example, it has been observed that the KMC algorithm takes an average time of 193 ms for 2000 data points while it takes 824 in FCMC algorithm for the same dataset size. Further, the KMC algorithm performs analysis tasks for 6000 data points in about 266 ms, while it takes 1693 in FCMC algorithm for the same dataset size. The FCMC algorithm has a higher performance overhead for the analysis tasks compared with the KMC algorithm, which is related the computation complexity of the FCMC algorithm compared with the KMC algorithm.

Regarding Intel SGX enclave memory usage for storing encrypted data, a dataset of 2000 encrypted data points consumes around 608 kilobytes of memory while the memory size increases in linear relation to the size of input dataset, as shown in Fig. 9.

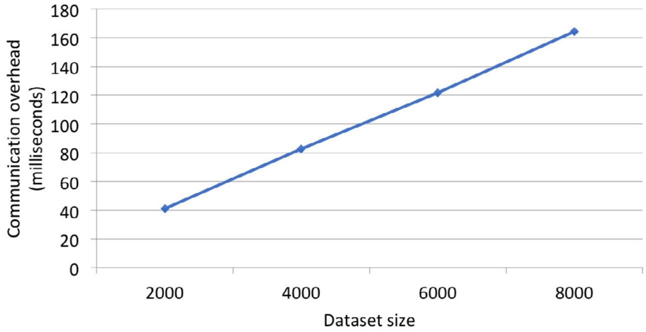


Fig. 10. Communication overheads for exchanging encrypted data with Intel SGX encrypted datasets of varying sizes.

Finally, one of the main obstacles in building SGX-based solutions for analytic models is the communication overhead, which is an essential component of analytic processes in which data are sent inside the SGX enclave through a secure established communication channel with third parties. Fig. 10 shows the approximate communication overhead between the remote IoT entity and the cloud server based on the size of the dataset, which provides a visible insight into the developed model's capabilities and limitations. For instance, it takes approximately 41 ms to transmit 2000 data points and approximately 82 ms to transmit 4000 data points. This shows a linear increase in the communication overhead with the size of input dataset, as shown in Fig. 10.

8. Conclusion

In this paper, a practical hybrid confidentiality-based analytic framework is based on Intel SGX. It relies on a hybrid-integrated model, including both software- and hardware-based solutions, to ensure the confidentiality and process sensitivity of outsourced data in the cloud environment. The developed framework aims to provide secure data-analytic services for IoT-enabled applications in various domains, such as smart grid and healthcare applications. The experimental evaluation shows a high level of analysis accuracy in a privacy-preserving manner, while indicating differences in execution times and processing overheads. The developed framework can be adapted efficiently for various analytical service applications, to take advantage of public cloud computing without compromising data privacy. Future research will focus on building more advanced analytical models, in order to overcome challenges such as communication and storage limitations, because of their complexity in both computational and analytical structure.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agrawal, R., Srikant, R., 2000. Privacy-preserving data mining. In: Chen, W., Naughton, J.F., Bernstein, P.A. (Eds.), Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, May 16-18, 2000, Dallas, Texas, USA. ACM, pp. 439–450.

Arnautov, S., Trach, B., Gregor, F., Knauth, T., Martin, A., Priebe, C., Lind, J., Muthukumaran, D., O'Keeffe, D., Stillwell, M., Goltzsche, D., Eyers, D.M., Kapitza, R., Pietzuch, P.R., Fetzer, C., 2016. SCONE: Secure linux containers with intel SGX. In: Keeton, K., Roscoe, T. (Eds.), 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016, Savannah, GA, USA, November 2-4, 2016. USENIX Association, pp. 689–703.

Atallah, M.J., Pantazopoulos, K.N., Rice, J.R., Spafford, E.H., 2001. Secure outsourcing of scientific computations. Adv. Comput. 54, 215–272.

Backes, M., Fiore, D., Reischuk, R.M., 2013. Verifiable delegation of computation on outsourced data. In: 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013. pp. 863–874.

Chadwick, D.W., Fatema, K., 2012. A privacy preserving authorisation system for the cloud. J. Comput. System Sci. 78 (5), 1359–1373.

Chen, G., Chen, S., Xiao, Y., Zhang, Y., Lin, Z., Lai, T., 2020. SgxPectre: Stealing intel secrets from SGX enclaves via speculative execution. IEEE Secur. Privacy 18 (3), 28–37.

Daemen, J., Rijmen, V., 2020. The Design of Rijndael - The Advanced Encryption Standard (AES), Second Edition. In: Information Security and Cryptography, Springer.

Doganay, M.C., Pedersen, T.B., Saygin, Y., Savas, E., Levi, A., 2008. Distributed privacy preserving k-means clustering with additive secret sharing. In: Proceedings of the 2008 International Workshop on Privacy and Anonymity in Information Society, PAIS 2008, Nantes, France, March 29, 2008. pp. 3–11.

Du, W., Zhan, J.Z., 2003. Using randomized response techniques for privacy-preserving data mining. In: Getoor, L., Senator, T.E., Domingos, P.M., Faloutsos, C. (Eds.), Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003. ACM, pp. 505–510.

Duan, Y., Canny, J.F., 2014. Practical distributed privacy-preserving data analysis at large scale. In: Gkoulalas-Divanis, A., Labbi, A. (Eds.), Large-Scale Data Analytics. Springer, pp. 219–252.

Flexera, 2019. Cloud computing trends: 2019 state of the cloud survey.

Franti, P., Virmajoki, O., 2006. Iterative shrinking method for clustering problems. Pattern Recognit. 39 (5), 761–775.

Gamal, T.E., 1985. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Trans. Inf. Theory 31 (4), 469–472.

Gentry, C., Halevi, S., 2011. Implementing gentry's fully-homomorphic encryption scheme. In: Paterson, K.G. (Ed.), Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings. In: Lecture Notes in Computer Science, vol. 6632, Springer, pp. 129–148.

Goldreich, O., 2005. Foundations of cryptography - A primer. Found. Trends Theor. Comput. Sci. 1 (1).

Grobauer, B., Walloschek, T., Stocker, E., 2010. Understanding cloud computing vulnerabilities. IEEE Secur. Privacy 9 (2), 50–57.

Hunt, T., Zhu, Z., Xu, Y., Peter, S., Witchel, E., 2018. Ryoan: A distributed sandbox for untrusted computation on secret data. ACM Trans. Comput. Syst. 35 (4), 13:1–13:32.

Inan, A., Kaya, S.V., Saygin, Y., Savas, E., Hintoglu, A.A., Levi, A., 2007. Privacy preserving clustering on horizontally partitioned data. Data Knowl. Eng. 63 (3), 646–666.

Langenberg, B., Pham, H., Steinwandt, R., 2020. Reducing the cost of implementing the advanced encryption standard as a quantum circuit. IEEE Trans. Quantum Eng. 1, 1–12.

Li, J., Huang, X., Li, J., Chen, X., Xiang, Y., 2014. Securely outsourcing attribute-based encryption with checkability. IEEE Trans. Parallel Distrib. Syst. 25 (8), 2201–2210.

McKeen, F., Alexandrovich, I., Berenzon, A., Rozas, C.V., Shafi, H., Shanbhogue, V., Savagaonkar, U.R., 2013. Innovative instructions and software model for isolated execution. In: HASP 2013, the Second Workshop on Hardware and Architectural Support for Security and Privacy, Tel-Aviv, Israel, June 23-24, 2013. p. 10.

Murdock, K., Oswald, D., Garcia, F.D., Bulck, J.V., Gruss, D., Piessens, F., 2020. Plundervolt: Software-based fault injection attacks against intel SGX. In: 2020 IEEE Symposium on Security and Privacy, SP 2020, San Francisco, CA, USA, May 18-21, 2020. IEEE, pp. 1466–1482.

Nechvatal, J., Barker, E., Bassham, L., Burr, W., Dworkin, M., Foti, J., Roback, E., 2001. Report on the development of the Advanced Encryption Standard (AES). J. Res. Natl. Inst. Stand. Technol. 106 (3), 511.

Ohrimenko, O., Schuster, F., Fournet, C., Mehta, A., Nowozin, S., Vaswani, K., Costa, M., 2016. Oblivious multi-party machine learning on trusted processors. In: Holz, T., Savage, S. (Eds.), 25th USENIX Security Symposium, USENIX Security 16, Austin, TX, USA, August 10-12, 2016. USENIX Association, pp. 619–636.

Oleksenko, O., Trach, B., Krahn, R., Silberstein, M., Fetzer, C., 2018. Varys: Protecting SGX enclaves from practical side-channel attacks. In: Gunawi, H.S., Reed, B. (Eds.), 2018 USENIX Annual Technical Conference, USENIX ATC 2018, Boston, MA, USA, July 11-13, 2018. USENIX Association, pp. 227–240.

Orenbach, M., Baumann, A., Silberstein, M., 2020. Autarky: closing controlled channels with self-paging enclaves. In: Bilas, A., Magoutis, K., Markatos, E.P., Kostic, D., Seltzer, M.I. (Eds.), EuroSys '20: Fifteenth EuroSys Conference 2020, Heraklion, Greece, April 27-30, 2020. ACM, pp. 7:1–7:16.

Oukili, S., Bri, S., 2017. High speed efficient advanced encryption standard implementation. In: 2017 International Symposium on Networks, Computers and Communications, ISNCC 2017, Marrakech, Morocco, May 16-18, 2017. IEEE, pp. 1–4.

Patel, S.J., Punjani, D., Jinwala, D.C., 2015. An efficient approach for privacy preserving distributed clustering in semi-honest model using elliptic curve cryptography. Int. J. Netw. Secur. 17 (3), 328–339.

Rao Rupanagudi, S., Vidya J, V., Bhat, V.G., Padmavathi, P., Darshan, G., Gurikar, S.K., Darshan, S., Sindhu, N., 2019. A further optimized mix column architecture design for the advanced encryption standard. In: 2019 11th International Conference on Knowledge and Smart Technology (KST). pp. 181–185.

Ren, K., Wang, C., Wang, Q., 2012. Security challenges for the public cloud. IEEE Internet Comput. 16 (1), 69–73.

Rivest, R.L., Shamir, A., Adleman, L.M., 1978. A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM 21 (2), 120–126.

Sasy, S., Gorbunov, S., Fletcher, C.W., 2018. Zerotrace : Oblivious memory primitives from intel SGX. In: 25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018. The Internet Society.

Schuster, F., Costa, M., Fournet, C., Gkantsidis, C., Peinado, M., Mainar-Ruiz, G., Russinovich, M., 2015. VC3: Trustworthy data analytics in the cloud using SGX. In: 2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015. IEEE Computer Society, pp. 38–54.

Sundareswaran, S., Squicciarini, A., Lin, D., 2012. Ensuring distributed accountability for data sharing in the cloud. IEEE Trans. Dependable Secure Comput. 9 (4), 556–568.

Wang, J., Zhao, Y., Jiang, S., Le, J., 2010. Providing privacy preserving in cloud computing. In: 3rd International Conference on Human System Interaction. pp. 472–475.

Zheng, W., Dave, A., Beekman, J.G., Popa, R.A., Gonzalez, J.E., Stoica, I., 2017. Opaque: An oblivious and encrypted distributed analytics platform. In: Akella, A., Howell, J. (Eds.), 14th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2017, Boston, MA, USA, March 27-29, 2017. USENIX Association, pp. 283–298.

Zhou, M., Mu, Y., Susilo, W., Au, M.H., Yan, J., 2011. Privacy-preserved access control for cloud computing. In: IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2011, Changsha, China, 16-18 November, 2011. pp. 83–90.

Zissis, D., Lekkas, D., 2012. Addressing cloud computing security issues. Future Gener. Comput. Syst. 28 (3), 583–592.