Secure Data Processing in the Cloud

Zoltán Ádám Mann[1], Eliot Salant[2], Mike Surridge[3], Dhouha Ayed[4], John Boyle[5], Maritta Heisel[1], Andreas Metzger[1], and Paul Mundt[6]

[1] University of Duisburg-Essen, Germany
[2] IBM Haifa Research Labs, Israel
[3] University of Southampton IT Innovation Centre, UK
[4] Thales Services, France
[5] Oxford Computer Consultants, UK
[6] Adaptant Solutions AG, Germany
http://www.restassuredh2020.eu/

Abstract. Data protection is a key issue in the adoption of cloud services. The project “RestAssured – Secure Data Processing in the Cloud,” financed by the European Union’s Horizon 2020 research and innovation programme, addresses the challenge of data protection in the cloud with a combination of innovative security solutions, data lifecycle management techniques, run-time adaptation, and automated risk management. This paper gives an overview about the project’s goals and current status.

Keywords: Cloud computing, Data protection, Privacy, Secure hardware enclaves, Sticky policies, Run-time adaptation, Automated risk management.

1 Project Objectives

Secure cloud computing is key for business success and end user adoption of federated and decentralized cloud services, and as such, is essential to stimulating the growth of the European Digital Single Market. And, while cloud-based data be kept secure, these data must also be made accessible to authorized users while ensuring privacy regulations such as the European Union’s General Data Protection Regulation (GDPR)[1].

The RestAssured project aims to provide solutions to specific technical concerns of data protection in the cloud through four main areas of innovation (see Fig. 1):

- Use of emerging hardware solutions such as Intel’s SGX to provide secure enclaves for data operations.
- Implementation of sticky policies which define data access, usage and storage rules.
- Run-time data protection assurance using self-adaptation and models@runtime.
- Automated risk management to automatically detect risks to data protection and rapidly determine the cost vs. benefits of alternative protection mechanisms.

[1] http://www.eugdpr.org/

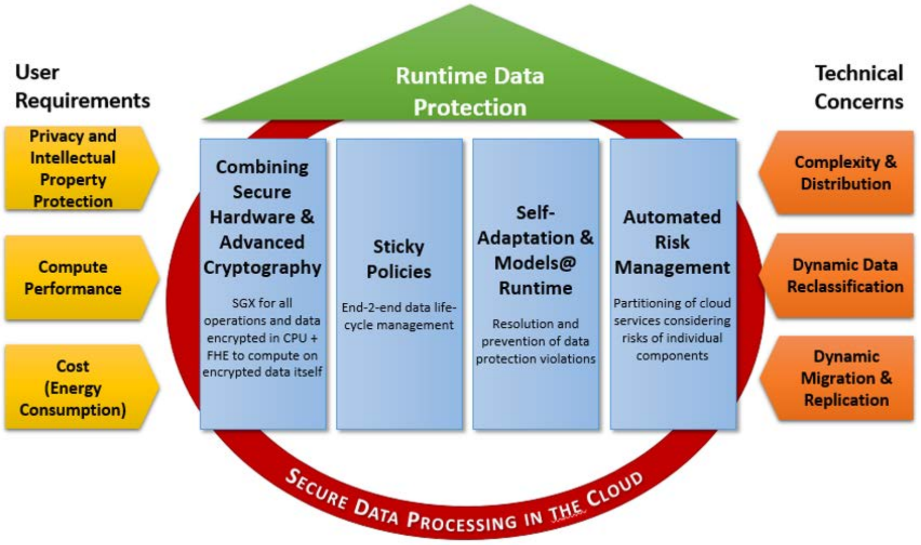
User Requirements
Privacy and Intellectual Property Protection
Compute Performance
Cost (Energy Consumption)
Runtime Data Protection
Combining Secure Hardware & Advanced Cryptography
SGX for all operations and data encrypted in CPU + FHE to compute on encrypted data itself
Sticky Policies
End-2-end data life-cycle management
Self-Adaptation & Models@ Runtime
Resolution and prevention of data protection violations
Automated Risk Management
Partitioning of cloud services considering risks of individual components
Technical Concerns
Complexity & Distribution
Dynamic Data Reclassification
Dynamic Migration & Replication
Secure Data Processing in the Cloud

Fig. 1. The RestAssured Pillars of Innovation

1.1 Secure Enclaves

Secure enclaves are offered by Intel's SGX (Software Guard Extensions) which is currently available in the marketplace, or AMD's SME (Secure Memory Encryption). While SGX and SME use different approaches, each with its advantages and disadvantages, the general idea is the same: a memory range is encrypted by the processor by a key which is generated at power-on, and not available to any running process. This means that all code and data within an enclave are protected from tamper and snooping, even by processes running at superuser level, or by dumping memory.

RestAssured is creating a toolkit which will significantly simplify the work required by a developer to set up and use an SGX enclave (such as remote attestation, sealing, secret passing), allowing developers to focus on the development of their business logic. Additionally, RestAssured is integrating the Open Source Opaque project into its SGX toolkit. Opaque is a Spark SQL engine that can work with encrypted data, leveraging Intel SGX to protect the computation [1]. Users can run SQL queries in a Spark shell, or program the queries in high-level Scala language. There is no need to develop SGX applications in C/C++ with the SGX SDK. However, Opaque has some design and implementation limitations, related to attestation and data key passing. By integrating with the RestAssured toolkit, we enable efficient attestation of Opaque enclaves, flexible data key passing and overall integration into the RestAssured platform.

1.2 Sticky Policies

Sticky policies for data define access rights on the data and, as their name suggests, "stick" to the data, following it as it migrates across the cloud. In this way, sticky policies allow for decentralized data lifecycle management; i.e. access control can be enforced by decision points across the cloud, without the need for a centralized enforcement entity. Sticky policies need not only support the rights of the data subject in accordance with GDPR requirements, but must also be able to support the security and privacy regulations which may be mandated by the enterprise which either owns or processes the data, as well as any regulations the data may be subject to, based on its physical storage location across the cloud.

1.3 Run-Time Data Protection Assurance

The flexibility and dynamism of the cloud poses a big challenge for data protection. The applicability of traditional security mechanisms designed to keep the system in a stable secure state is limited. In particular, security-by-design methodologies are not sufficient, due to uncertainty at design time as to how the cloud and privacy requirements may evolve and change at run time.

To cope with continuously changing data protection requirements in a continuously changing cloud environment, we apply methods from the field of self-adaptive systems [2,3]. This way the system can adapt to changes in both the cloud and the data protection requirements, ensuring that requirements are met in the presence of changes, with minimal impact on performance and costs. Adaptations may be made either fully automatically, or after approval from a human operator.

To make sound adaptation decisions at run time, one needs a model of the system, its requirements and environment. The model must be available at run time to enable online reasoning, hence it is called model@runtime [4]. Data protection concerns relate to all layers of the cloud stack, including secure hardware capabilities, co-location of different tenants on the same server, encryption of communication between application components, and data anonymization. All must be captured by the model@runtime. By monitoring the state of the system and its environment, updating the model, and comparing observed behavior to expected behavior, violations of data protection policies can be detected. If a violation is detected, further reasoning using the model@runtime can be used to automatically find and execute an appropriate adaptation action. This way, data protection issues can be mitigated or prevented automatically.

1.4 Automated Risk Management

Under the GDPR, personal data controllers and processors must assess risks to personal data, and employ security measures to appropriately manage identified risks, throughout the life of the system(s) storing and using the data. Moreover, the data controller is responsible for proving the systems and processes used comply with the GDPR. It is no longer enough to implement recommended security measures based on a ‘generic’ risk analysis – one must analyze risks specific to each situation, design systems and processes to address risks to privacy, and continuously review and update the risk analysis and security measures as either the system or the threat landscape evolves.

The requirement for continuous and auditable management of risks is especially difficult in cloud-based applications, which may be subject to automatic adaptation at any time. One of the main goals of RestAssured is to provide the means to trace how such changes affect the level of risk, and where changes are made specifically to manage risks, e.g. by allocating sensitive processes to a secure enclave, or by introducing advanced encryption to block new risks when migrating workloads. The goal is to provide technologies that help data controllers to analyze risks and trace the measures used to address risks, making it much easier to comply with GDPR when using cloud-based applications. To achieve this, RestAssured integrates and extends two innovative approaches to information risk analysis:

- the Cloud System Analysis Pattern methodology [5] developed by University of Duisburg-Essen in the CloudDAT project to help stakeholders identify socio-technical assets and carry out a risk analysis specifically focusing on cloud applications;
- a procedure devised by IT Innovation [6,7] in the SERSCIS, OPTET and UK ASSURED projects that automates risk identification and analysis based on a description of a system in terms of its assets, and supports the selection of risk management measures.

The use of an asset-based risk analysis approach supports compliance with information risk management standards like ISO 27001 [8], while the use of automation based on machine reasoning makes it possible to perform this analysis on a continuous basis in the loop of autonomic cloud application and infrastructure management processes.

2 Project Current State and Summary of Results

Although the project is still in its first year, good progress has been made towards having a prototype implementation running two real-world use cases by project month 18.

The first use case, highlighting social care services, shows how volunteer healthcare workers can be matched with suitable healthcare patients in a secure environment, preserving the data access rights specified by both parties. Additionally, this use case demonstrates how RestAssured can enforce the security and privacy requirements for a workflow specifying the generation of summary reports by a third party only allowing them access to anonymized data. Pivotal to this use case is the ability to integrate Opaque into the RestAssured environment to support the database queries required to match caregivers with patients, as well as sticky policy enforcement across the whole workflow of the data.

An additional use case demonstrates a “pay-as-you-drive” scenario – where a driver’s insurance rates depend on their monitored driving behavior. In this scenario, an application at the network edge (e.g. Connected Car) enables the data subject (driver) to identify and limit the transfer of personally identifiable information to the service provider for providing an agreed upon service (e.g. usage-based car insurance). This can be attained through the application of policies that match the intent of the data subject to the data, while also enabling the data subject to apply data minimization to certain data (e.g. sensitive data the data subject is not comfortable sharing, or data deemed not to be relevant for the purpose of service contextualization) prior to its transfer to the service provider. Data subjects are able to opt-in/out of secondary/tertiary processing of the data beyond the original agreed-upon purpose and the transfer to third party organisations, as per their rights under the GDPR.

As in the previous use case, sticky policies, secure enclaves, and the RestAssured toolkit play a central role in simplifying what is required by developers to implement and deploy SGX-based applications.

From a technical perspective, a first draft of a prototype architecture for RestAssured has been developed. This architectural blueprint defines the functionality of all major system components, and defines the high-level interfaces between them.

Acknowledgement. The research leading to these results has received funding from the European Community's Horizon 2020 research and innovation programme under grant agreement n° 731678.

References

1. Zheng, W., Dave, A., Beekman, J.G., Popa, R.A., Gonzalez, J.E. and Stoica, I.: Opaque: An Oblivious and Encrypted Distributed Analytics Platform. In Proc. of the 14th USENIX symposium on Networked Systems Design and Implementation (NSDI'17), pp. 283-298, USENIX Assoc (2017).
2. Mann, Z. Á., Metzger, A.: Optimized Cloud Deployment of Multi-tenant Software Considering Data Protection Concerns. In: Proc. of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2017), pp. 609-618, IEEE Press (2017).
3. Dräxler, S., Karl, H., Mann, Z. Á.: Joint optimization of scaling and placement of virtual network services. In: Proc. of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2017), pp. 365-370, IEEE Press (2017).
4. Schoenen, S., Mann, Z. Á., Metzger, A.: Using risk patterns to identify violations of data protection policies in cloud systems. In: 13th International Workshop on Engineering Service-Oriented Applications and Cloud Services (WESOACS) (2017).
5. Beckers, K., Côté, I., Goeke, L., Güler, S. and Heisel, M.,: A structured method for security requirements elicitation concerning the cloud computing domain. In International Journal of Secure Software Engineering (IJSSE), 5(2), pp.20-43 (2014).
6. Surridge, M., Nasser, B., Chen, X., Chakravarthy, A., Melas, P.: Run-Time Risk Management in Adaptive ICT Systems. In: 8th International Conference on Availability, Reliability and Security (ARES), pp. 102-110, IEEE (2013).
7. Chakravarthy, A., Wiegand, S., Chen, X., Nasser, B. and Surridge, M.: Trustworthy Systems Design using Semantic Risk Modelling. In: Proc. of the 1st International Conference on Cyber Security for Sustainable Society, pp. 49-81, Digital Economy Sustainable Society Network (2015).
8. ISO/IEC 27001:2013. Information technology – Security Techniques – Information security management systems – Requirements, International Organization for Standardization (2013).