A GDPR-compliant approach to real-time processing of sensitive data

Luigi Sgaglione[1] and Giovanni Mazzeo[1]

[1] University of Naples "Parthenope", Naples, Italy
{luigi.sgaglione;giovanni.mazzeo}@uniparthenope.it

Abstract. Cyber-attacks represent a serious threat to public authorities and their agencies are an attractive target for hackers. The public sector as a whole collects lots of data on its citizens, but that data is often kept on vulnerable systems. Especially for Local Public Administrations (LPAs), protection against cyber-attacks is an extremely relevant issue due to outdated technologies and budget constraints. Furthermore, the General Data Protection Regulation (GDPR) poses many constraints/limitations on the data usage when "special type of data" is processed. In this paper the approach of the EU project COMPACT (H2020) is presented and the solutions used to guarantee the data privacy during the real time monitoring performed by the COMPACT security tools are highlighted.

Keywords: real time processing, SIEM, SOC, data privacy, homomorphic encryption.

1 Introduction

The advent of the Internet has been opening new opportunities for Public Administrations (PAs) to improve their efficiency while providing better services to citizens via an ever larger set of specialized network applications, including e-government, e-health, and more. This is at the heart of a European wide eGovernment action plan, whose latest update covers the years 2016 to 2020 and which also mentions the importance of trustworthiness and security as a key guiding value. Indeed, as a potential channel of accessing personal information, these specialized applications also expose the public sector to new risks.

The cybersecurity landscape is changing, and Local Public Administrations (LPAs) and Critical Infrastructures (CIs) are rapidly becoming an attractive target for cybercriminals [1,2,3,4,5], who might access some sets of personal data or gain control over smartly operated city resources through LPAs/CIs infrastructures. The consequences of cyber-threats have the potential to be considerable causing business interruptions, data losses, and thefts of intellectual property, significantly impacting both individuals and organizations.

It is claimed that cyber-threats are the most significant and rising risk that public sector organizations are facing. Reports demonstrates that nearly 40% of malware

attacks and in general cyber threats to which public bodies have been subject [6] are against public sector organizations [1], i.e. more than sectors (e.g. finance) which have traditionally been thought of as top targets. The interconnection of operational environment systems, used by the public bodies in ever growing scale, exacerbates the problem, especially as malware distribution periods (both fixed and mobile) are becoming increasingly short [7]. The increase in severity of cyber-attacks coincides with a boom in the different types of connected devices, as well as with a huge expansion in virtualization and public clouds.

In particular, Information Commissioner's Office (ICO) reports that: "In a change to the previous quarter, the second most prevalent sector in Q4 (January to March 2016)1 was local government. The number of data security incidents in this sector increased by 34% compared to the previous quarter (from 32% in Q3 to 43% in Q4).

Coupled with the overall decrease in data security incidents during Q4, this means that the percentage of total incidents suffered by the local government sector has also increased, from 6% in Q3 to 10% in Q4." [8].

Therefore, LPAs need to understand the cyber risks to which they are exposed and take proper actions to protect their infrastructures from cyber disruptions, to safeguard citizen's and enterprises' information they manage. The DBIR 2016 report [9] provides the number of security incidents by victim industry and organization size (2015 dataset). The category "Public Industry" – which refers to PA organizations – is by far the most targeted, with 47000+ attacks out of a total of about 64000. The report also shows the distribution of incidents by patterns: the vast majority of incidents in the public sector can be rooted to: 1) miscellaneous errors (24%), 2) privilege misuse (22%), 3) stolen assets (20%), and 4) crimeware (16%).

The issues that have been identified and that hamper the ability of PA organizations of improving their cyber security level, most notably are:

1. **Lack of standardized data classification** – 45% of public sector respondents do not use standardized data classification techniques/procedures. As a consequence, LPAs run a higher risk of accidentally exposing private data in their rush to comply with emerging regulations – both at the national and at the EU level – promoting transparency of the Public Sector. Also, only 12% stated that they used standardized policies and that they proactively verify and enforce those policies.
2. **Lack of effective Non-Disclosure Agreements** (NDAs) – 40% of public sector organizations still rely on paper-based NDAs, and use them inconsistently. This amplifies risks related to the human factor, which is already one of the biggest, since malicious or disgruntled personnel with access to important information assets can be a significant threat to the security of those assets.
3. **Lack of plans for responding to security breaches and for disaster recovery** – 36% of public sector organizations do not have a plan for responding to security breaches, and only 10% of public sector organizations test for the worst-case scenario. 34% of public sector organizations do not have budgeted disaster recovery plans. These are major impairments to contain the damage, since when a security incident or a disaster occur, proper and timely action is key.

4. **Lack of uniformly enforced security policies** – 33% of public sector organizations do not have uniformly enforced security policies (this means limited application - if not complete lack - of a consistent security policy throughout the whole organization.). This condition hinders their ability to comply with regulations, such as the European Union Data Protection Directive (EUDPD).
5. **Lack of adequate policies and practices for data disposal** – 76% of public sector organizations do not have adequate policies and practices for secure and reliable data disposal. In particular, only 16% of public sector organizations have written policies that require destruction records to be actually collected, practiced, and audited. The enforcement of strong policies to govern the proper disposal of electronic and paper records - based on sound technical and organizational guidelines and best practices - is the prerequisite for protecting private data from unauthorized disclosure.
6. **Lack of effective access control mechanisms** – 20% of public sector organizations do not use roles to manage access, and more than 26% of public sector organizations have no official procedure for terminated or reassigned employees. This create vulnerabilities, since it allows inappropriate access to resources.
7. **Large set of legacy unmaintained and undocumented systems** representing an attack surface of unknown dimension.
8. **Inappropriate management of security updates** (patches), as well as usage of out of date software in computers, mobile devices and central servers.
9. **Limited capacity, and motivation, of LPAs personnel** in detecting and reporting cyber-attacks. This is due to a number of interconnected factors including (i) the aging of the LPAs workforce, (ii) its limited technological skills and (iii) the lack of acknowledgment of employees' achievements. This makes the PA workforces less responsive to the traditional educational measures (like classroom training).

It is clear that innovative cyber security tools are needed in order to guarantee the protection of LPAs. In addition, these tools must to deal with:

(1) Limited resources in terms of both economic and structural
(2) Strong privacy requirements coming from the recent adoption of the General Data Protection Regulation (GDPR) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

2 Backgroud – Homomorphic encryption

2.1 Homomorphic encryption

Homomorphic Encryption is a recent cryptographic method which allows to perform computation on encrypted data without decrypting it. This way, the confidential data can be protected not only during the storage and exchange/transfer but also during the

processing. Avoiding intrusions from semi-honest or malicious cloud providers when outsourcing data processing to the Cloud is crucial for the case of sensitive data that are about to be processed in frames of the COMPACT solution.

The first HE algorithms, i.e., Partially Homomorphic Encryption (PHE) [14] [15], had the ability to carry out just one type of operations (e.g., addition, or multiplication). Clearly, the limitation in the type of executable computations hampered the usage of HE in practical contexts. Gentry et al. [16] provided the first implementation of a Fully Homomorphic Encryption (FHE) scheme. Gentry's algorithm allows the execution of an arbitrary number of additions and multiplications over encrypted data. The security of the system is based on the noise introduced into the ciphered text. When the noise reaches some maximum amount, the ciphertext becomes undecryptable. This solution was very costly in terms of performance. It highly affects CPU and memory resources.

An attempt to simplify the method has been provided by Van Dijk et al. [17] who proposed a FHE i.e., Somewhat Homomorphic Encryption (SHE) over the integers. The price to pay with SHE is given by the limited number of mathematical operations that can be performed. However, in many real-world applications (e.g., medical, financial) this seems reasonable since – as Naehrig et al. [9] analysis reports – most of the evaluations required, i.e., one-time statistical functions, fits well with SHE constraints.

Among the aims of COMPACT are to adopt Fully Homomorphic Encryption (FHE) Schemes capable of performing any arbitrary function in an homomorphic way and to mitigate performance overheads introduced by Homomorphic computation, using recent dedicated compilation and parallelism techniques and mechanisms.

3 The COMPACT project

COMPACT's overarching objective is to enable LPAs to become the main actors of their own cyber-resilience improvement process, by providing them with effective tools and services for removing security bottlenecks. This can be broken down into five finer-grain objectives:

- Objective #1 - Making the PA personnel aware of the basic cyber security threats they are exposed to.
- Objective #2 - Improving the skills – both technical and behavioral – of the PA personnel via innovative training techniques that are well received by the (non IT-expert) workforce.
- Objective #3 - Providing protection tools against basic cyber security threats, i.e. those with a higher impact on LPAs. These include [10,11,12]: phishing, ransomware, Bring Your Own Device (BYOD), jailbreaking the cloud, cross-site scripting, code (particularly SQL) injection, and more.
- Objective #4 - Creating a LPAs level information hub, for favouring reliable and timely exchange of information among LPAs on cyber security guidelines and best practices, as well as on Indicators of Compromise (IoC).

- Objective #5 - Creating a link between COMPACT LPAs level information hub and major EU level initiatives, for supporting LPAs to improve cyber-resilience in a complex European context.

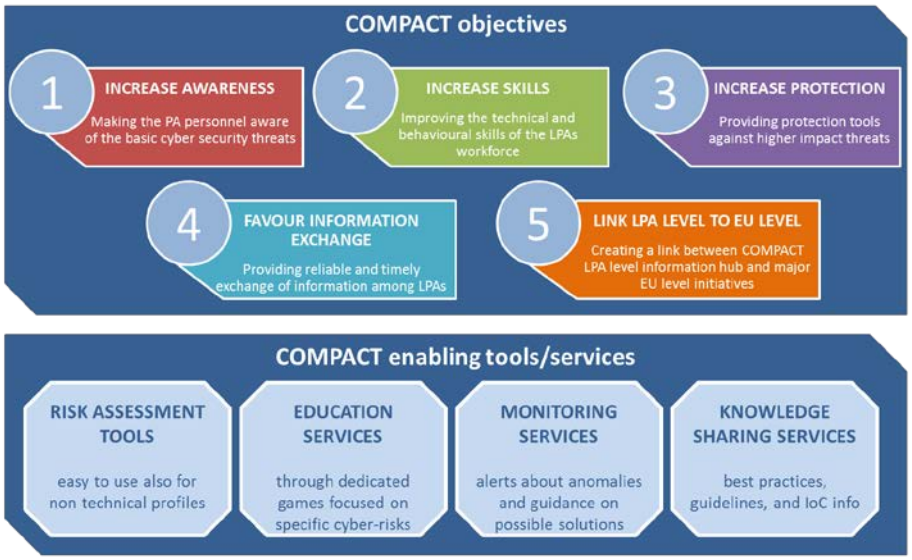

Fig. 1. COMPACT objectives

To achieve its objectives, COMPACT will develop four types of tools/services (**Fig. 1**), which include:

1. Risk assessment tools - Tailored to the LPAs context that will allow LPAs to evaluate and monitor their exposure to the most relevant (i.e. with the highest impact) cyber treats. They will enable LPAs to prioritize the adoption of preventive and reactive countermeasures, for maximum efficiency of resource usage for cyber protection purposes.
2. Education services - Through dedicated game-based training, focused not only on specific cyber-threats but also on psychological and behavioral factors, to maximize the effectiveness of the learning experience, while also containing the training time.
3. Monitoring services (SOC) - That continuously process events related to the status of the infrastructure and correlate them with information from threat intelligence feeds to timely spot anomalies and also suggest recovery actions that can be implemented.
4. Knowledge Sharing services – These will include best practices and guidelines, focused on the specific needs of LPAs, that can be easily adopted to quickly increase the cyber security level of the organization. Just as importantly, they are also used (i) at the Member States level as an input for the activity of national cybersecurity stakeholders (like national CERTs5) and

(ii) at the EU level as an input for European boards, agencies, and initiatives (like ENISA and the CSIRT [13] network foreseen in the NIS directive)

4 COMPACT monitoring service

The Security Operations Centre (SOC) provides, throughout advanced Security Information and Event Management (SIEM), the real-time monitoring capability of the organization. SOC platform is an integrated technology platform that allows for accurate, timely and trustworthy detection and diagnosis of security attacks, combining information from physical and logical event sources. The platform has been implemented in a distributed loosely interoperating architecture, where components depend on each other to the least extent practicable.

The SOC is implemented as a distributed architecture that enables: i) collection of security-relevant data from a variety of data feeds; ii) correlation of events and context information, via combined use of stream and batch processing; and iii) production and secure storage of incident-related evidence.

The event sources for SOC platform can be physical or logical alike. Physical event sources include physical systems that are existing in the buildings, like video surveillance system, physical access control system, fire alarm system, other physical security systems, or automation and building management systems, for example. Logical security systems can be defined to consist of software safeguards for an organization's systems, including user identification and password access, authenticating, access rights and authority levels.

SOC platform has the capability to combine event information from multiple event sources and to make sophisticated diagnosis based on the received information. As the outcome of the analysis performed by the SOC platform, the end user will receive ranked alerts and forensic evidences.

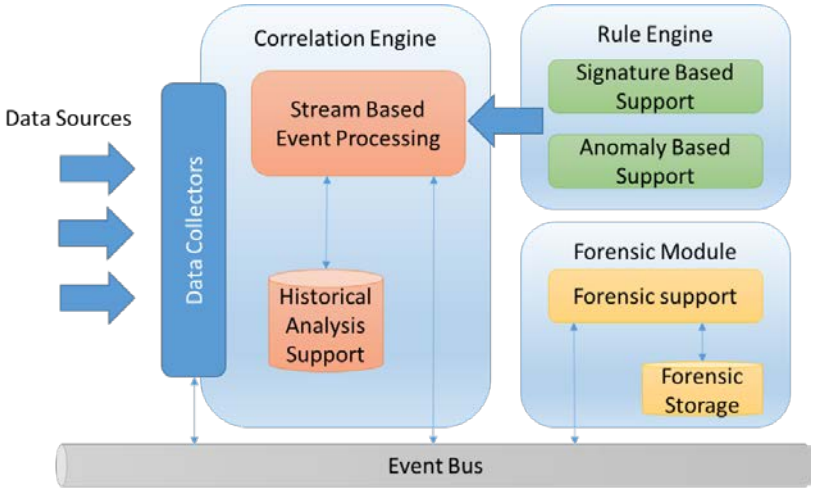

Fig. 2. SOC architecture

An architecture of the current solution is reported in **Fig. 2**. SOC platform consists of the following main components:

- Correlation Engine:

The Correlation Engine is the component in charge of the event diagnosis process. It operates by correlating a huge amount of security relevant events/information from the physical and the electronic domain in real-time, through Complex Event Processing (CEP) techniques and stream processing computing technologies.
The attack diagnosis process is driven by correlation rules that aggregate the parameters of attack symptoms, such as the attack type, the target component and the temporal proximity. Alerts are generated only when the correlation among such symptoms indicates a potential attack, thus exhibiting low false positive rates and improved detection capability w.r.t. single probes.

- Rule Engine:

The Rule Engine provides the logical rules to be followed for the Correlation Engine. The Rule Engine includes two main components, Signature Based Support and Anomaly Based Support.

- Forensic Module:

The Forensic Module provides a set of services that enables the end user (SOC operator) to trace from an event to the log data from which it was identified. The module will ensure that the events and their associated logs are stored in a forensically sound manner. It will support processes that ensure, to the greatest extent possible, that the event data will be acceptable as evidence.

In terms of data collection, the prototype is equipped with a number of adapters, for receiving events from a wide variety of Commercial Off The Shelf (COTS) products for logical and physical security monitoring. In terms of data processing, the prototype enables: 1) pre-processing of data at the edge of the system and 2) stream and batch processing in the core of the system. The business logic that drives the correlation process can be easily customized by means of a user-friendly graphical interface.
The SIEM is the main component of the SOC systems and includes:

- A runtime engine to allow the distributed streaming dataflow
- Two data processing APIs, one for the Stream Processing and one for the Batch Processing
- Three class of libraries:
    1. Complex Event Processing (CEP) to detect event patterns in an endless stream of events. Event processing combines data from multiple sources to infer events or patterns in order to highlight specific situations. The goal of complex event processing is to identify meaningful events (such as threats) and respond to them as quickly as possible. This real time elaboration can be based on a time window or event-driven approach.

2. Machine Learning that gives SIEM the ability to learn without being explicitly programmed. It requires the use of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs.
3. Homomorphic Data Processing to allow the processing of homomorphic encrypted data without decrypting them

The communication between the SOC component is provided by a Publish Subscribe communication channel: it is in charge of delivering the data and messages between data sources, SIEM GUI and SIEM Core.

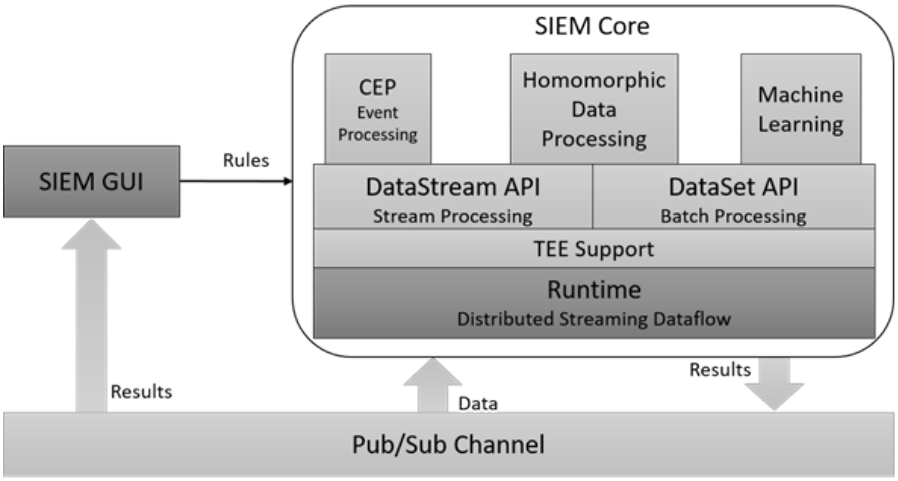

Fig. 3. SIEM components

Even a SOC prototype is already available; it will be evolved to meet the COMPACT requirements along several dimensions.

The first development will regard the improvement and adaptation of the SOC data collection to the data that must be acquired during the LPA monitoring. Many data collection features are already available in the current SOC prototype and these will be adapted to be compliant with the LPA environments; others will be developed to meet specific requirements likes the acquisition of information from the Windows Management Instrumentation tool and from other security tools (Nagios, Sophos, etc.).

The second improvement will be related to the implementation of the Data Management and Policy Enforcement component (DMPE). This component will be integrated in each data collection tool in order to fulfill the privacy requirements imposed by the LPA (to be compliant with the GDPR). In particular, the DMPE will be in charge of applying the most appropriate techniques needed to meet the privacy requirements,

such as anonymization and pseudo anonymization to remove special categories of data or homomorphic encryption to hide data and process it in an encrypted form.
The third improvement is related to the technology update of the current correlation and processing features of the SOC, by exploiting a best of breed selection of Open Source technologies for CEP, machine learning, and data mining.
The fourth improvement will be related to the implementation of specific correlation operators (CEP operators) able to process the homomorphically encrypted data without decrypting it.
Finally, the SOC graphical user interface will be developed/adapted in order to meet the guidelines defined by the COMPACT consortium and to be integrated with the COMPACT unified dashboard.

5 Conclusions

In this paper, a brief overview of the COMPACT approach used for the implementation of a Security Monitoring Center tailored to LPAs has been presented. The paper also highlights how this component will guarantee the privacy of sensitive data during the processing phase.

6 ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under grant agreements No 74071 (COMPACT)

References

1. Time to face up to cyber risk, http://www.publicfinance.co.uk/opinion/2016/03/time-face-cyber-risk, last accessed 2018/04/09.
2. Coppolino, L & D'Antonio, Salvatore & Romano, Luigi. (2014). Exposing vulnerabilities in electric power grids: An experimental approach. International Journal of Critical Infrastructure Protection. 7. 10.1016/j.ijcip.2014.01.003.
3. Coppolino, L & D'Antonio, Salvatore & Formicola, Valerio & Romano, Luigi. (2013). Enhancing SIEM Technology to Protect Critical Infrastructures. 10-21. 10.1007/978-3-642-41485-5_2.
4. Coppolino, L & D'Antonio, Salvatore & Formicola, Valerio & Romano, Luigi. (2011). Integration of a System for Critical Infrastructure Protection with the OSSIM SIEM Platform: A dam case study. 6894. 199-212. 10.1007/978-3-642-24270-0_15.
5. D'Antonio, Salvatore & Coppolino, L & Elia, Ivano & Formicola, Valerio. (2011). Security issues of a phasor data concentrator for smart grid infrastructure. 10.1145/1978582.1978584.
6. Data Breach Investigations Report (DBIR), http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/, last accessed 2018/04/09.

7. CCN-CERT, Threats and Risk Analysis in Industrial Control Systems (ICS), Report IA-04/16, Centro Criptologico Nacional: Madrid, 28 Jan 2016 (in Spanish), https://www.ccn-cert.cni.es/informes/informes-ccncert-publicos/1381-ccn-cert-ia-04-16-amenazas-y-analisis-de-riesgos-en-sistemas-de-control-industrial-ics/file.html, last accessed 2018/04/09.
8. Data security incident trends, https://ico.org.uk/action-weve-taken/data-security-incident-trends/, last accessed 2018/04/09.
9. A. Gajli, Time to face up to cyber risk, Public Finance, 31 March 2016, http://www.publicfinance.co.uk/opinion/2016/03/time-face-cyber-risk, last accessed 2018/04/09.
10. “3 Basic Cyber Security Threats To Be Aware Of That People Still Get Wrong”, http://blog.scstechsolutions.co.uk/3-basic-cyber-security-threats/, last accessed 2018/04/09.
11. “Biggest cybersecurity threats in 2016”, http://www.cnbc.com/2015/12/28/biggest-cybersecurity-threats-in-2016.html, last accessed 2018/04/09.
12. “Top 7 Cyberthreats to Watch Out for in 2015-2016”, Kaspersky Lab
13. "Computer Security and Incident Response Teams network" https://www.enisa.europa.eu/topics/national-csirt-network
14. T. El Gamal, A public key cryptosystem and a signature scheme based on discrete logarithms, in: Proceedings of CRYPTO 84 on Advances in Cryptology, Springer-Verlag New York, Inc., New York, NY, USA, 1985, pp. 10–18. URL http://dl.acm.org/citation.cfm?id=19478.19480
15. P. Paillier, Public-Key Cryptosystems Based on Composite Degree Residuosity Classes, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999, pp. 223–238. doi:10.1007/3-540-48910-X_16. URL http://dx.doi.org/10.1007/3-540-48910-X 16
16. C. Gentry, Fully homomorphic encryption using ideal lattices, in: Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing, STOC ’09, ACM, New York, NY, USA, 2009, pp. 169–178. doi:10.1145/1536414.1536440. URL http://doi.acm.org/10.1145/1536414.1536440
17. M. van Dijk, C. Gentry, S. Halevi, V. Vaikuntanathan, Fully homomorphic encryption over the integers, Cryptology ePrint Archive, Report 2009/616, http://eprint.iacr.org/2009/616 (2009).
18. M. Naehrig, K. Lauter, V. Vaikuntanathan, Can homomorphic encryption be practical?, in: Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop, CCSW ’11, ACM, New York, NY, USA, 2011, pp. 113–124. doi:10.1145/2046660.2046682. URL http://doi.acm.org/10.1145/2046660.2046682