Towards a Systematic Model-driven Approach for the Detection of Web Threats and Use Cases

Simona Bernardi[1], Raúl Piracés Alastuey[2], Alejandro Solanas Bonilla[2], and Raquel Trillo-Lado[2]

[1] Centro Universitario de la Defensa, Zaragoza, Spain, simonab@unizar.es
[2] Universidad de Zaragoza, Spain, raul.piraces@gmail.com, 647647@unizar.es, raqueltl@unizar.es

Abstract. The increasing use of Web Information System has made them an attractive target for attackers. Herein, we present firsts results and current work on a method for improving the security of such systems, which is based on Model-Driven Engineering and Process Mining.

Introduction. The Web has become a popular communication and information exchange channel, not only for people but also for different types of systems. Thus, for example, while previously cyber physical systems, such as the electrical networks, were isolated; nowadays, they are usually interconnected via information infrastructures where the Web is used. For example, the company Iberdrola Distribución Eléctrica offers its clients a service to consult their electrical consumptions via Web applications[3]. The increasing use of Web Information System has made them an attractive target for attackers. According to the last Symantec report published in April 2017 [2] *"Web attacks are still a big problem, with an average of more than 229,000 being detected every single day in 2016"*. Besides, the same report indicates that *"More than three-quarters (76 percent) of scanned websites in 2016 contained vulnerabilities, nine percent of which were deemed critical"*. So, improving the security of Web Information Systems in order to detect new threats and vulnerabilities is relevant, in particular in the context of critical infrastructures such as energy networks.

Approach overview. Recently, we have proposed a new method based on Model-Driven Engineering and Process Mining techniques for improving the security of Web Information Systems [1]. Our proposal consists of five main steps:

- *Step 1.* Specification of the expected system behavior by means of the Unified Modeling Language (UML) [3].
- *Step 2.* Automatic generation of a Petri net model from the UML-based specification by means of the DICE-tools [4]. This model formally specifies the expected system behavior and it is named *normative model*.
- *Step 3.* Control and monitoring of the Web Information System to get data logs that are evidences of the operative (or real) behavior of the system.

[3] https://www.iberdroladistribucionelectrica.com/consumidores/inicio.html

- *Step 4.* Pre-processing of the data logs to transform them into *event logs* for process mining.
- *Step 5.* Use of process mining techniques for the identification of deviations between the normative model and the operative behavior. The deviations are analyzed to determine if they are potential threats or new trends of use (new use cases) or missed use cases not considered when the normative model was specified.

When a potential threat is detected, new measurements to mitigate or remove the risk of its materialization are considered and deployed. On the other hand, when a new use case is detected it is analyzed to improve the services provided to the users (e.g., to offer customized services to the clients or to improve the usability of the Web system). Missed use cases are used to enhance the initial UML specifications and improve the performance of the method proposed.

The method was used to study the SID Digital Library[4] by considering its logs during the last seven years. Very promising results, that demonstrate the feasibility of the proposal, were achieved: new trends of usage were identified and threats, previously not detected, were discovered [1].

On-going work. To improve the approach, we are currently tackling several open issues such as: 1) define new heuristics to get event logs that enable the analysis on different levels of granularity; 2) develop plugins to automatize the method and enable the analysis of logs on-line; 3) create a library of attack patterns for testing purposes; and 4) apply the method to new case studies[5].

Acknowledgment. This work has been funded by the projects: "Desarrollo de técnicas de detección de ciberataques en sistemas de información mediante minería de procesos" [UZ-CUD2016-TEC-06], "Ciber-resilient critical infrastructures: Exploiting process mining techniques for security-by-design" [CyCriSec-TIN2014-58457-R], and "Developing Data-Intensive Cloud Applications with Iterative Quality Enhancements" [DICE-H2020-644869].

References

1. S. Bernardi, R. Piracés-Alastuey, R. Trillo-Lado, Using Process Mining and Model-driven Engineering to Enhance Security of Web Information Systems, 2nd Int. Workshop on Safety & Security aSSurance for Critical Infrastructures Protection (S4CIP), 29th April, 2017, Paris (France).
2. Symantec Corp. Global Internet Security Threat Report, vol. 22, April 2017.
3. Object Management Group. Unified Modeling Language (UML), v2.5, June 2015.
4. A. Gómez, C. Joubert, J. Merseguer, A Tool for Assessing Performance Requirements of Data-Intensive Applications, pp. 159–169, XXIV National Conference of Concurrency and Distributed Systems, 2016, ISBN: 978-84-16478-90-3.

[4] SID Digital Library: `http://sid.cps.unizar.es/BiD`

[5] The social network *Yarning*: `https://www.yarning.es` and the Content Management System *e-ditor*: `http://www.e-ditor.es`.