Active Vulnerability Assessment Tool

Margus Välja, MSc[1]; Ruben ten Hove, MSc[2]
[1]KTH, [2]TNO
Contact: [1]margusv@kth.se, [2]ruben.tenhove@tno.nl

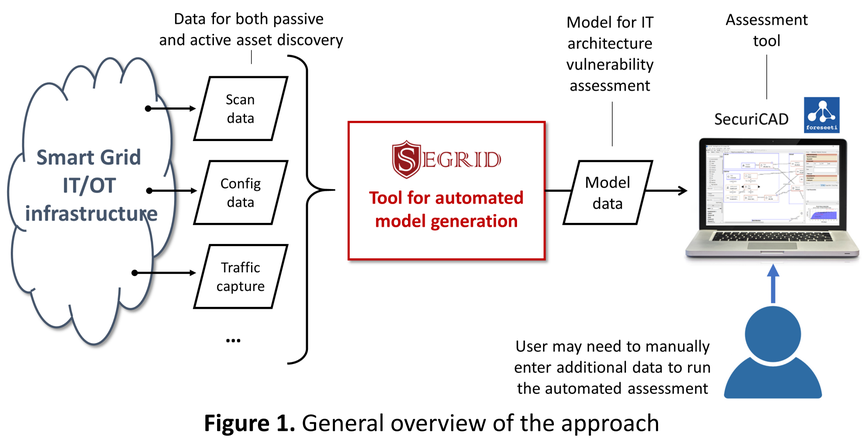

Figure 1. General overview of the approach

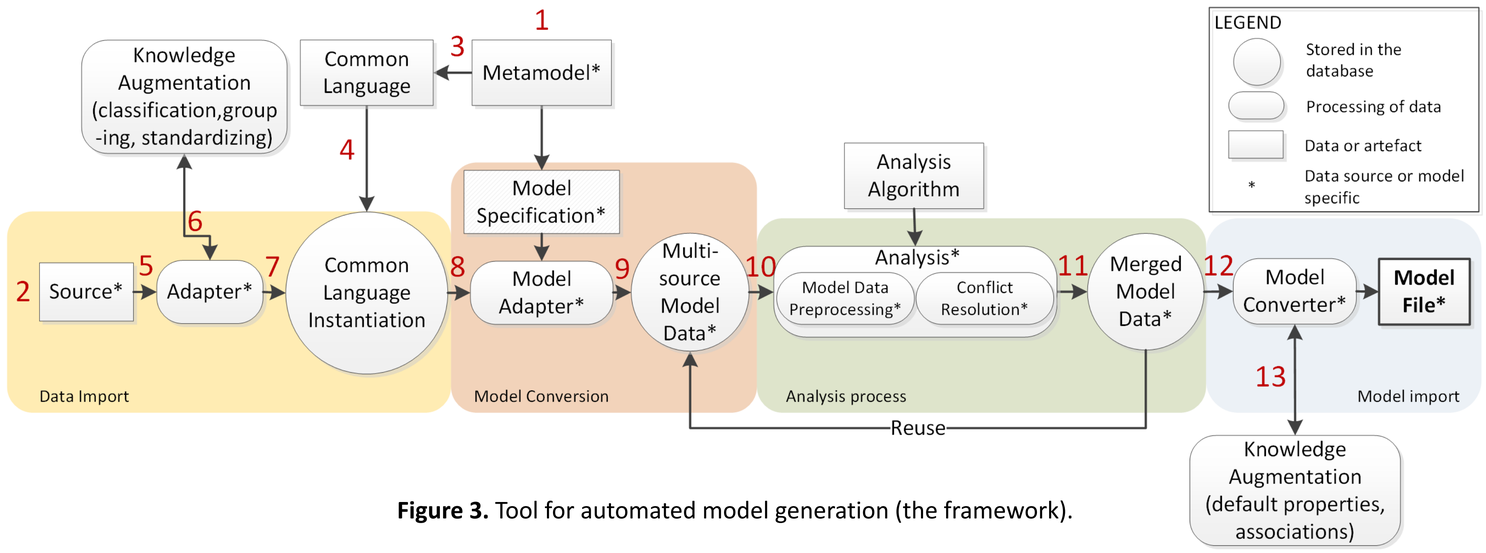

Figure 3. Tool for automated model generation (the framework).

Introduction

Modeling systems and software on (smart grid) infrastructure level is a complex, time consuming, and error prone task. Even more so in the field of cyber security. However, many of the ICT systems within the smart grid (both in IT and OT) produce information that can be used to automate (part of) the modeling. A problem is that these systems tend to be heterogeneous, and the data that can be extracted from them as well. These heterogeneous sources can provide data on different levels of granularity, and different formats and accuracy. To merge the heterogeneous data, the problems and conflicts in the data need to be resolved. A smart method is needed for merging, that allows using the knowledge we have about the data sources, the data itself and external data enrichment sources.

We propose a framework to automate the cyber security model creation that aims to solve the heterogeneity problems. The data in the framework is divided into claims, where claim means a statement a data source makes about some ICT object. The framework stores the data in a hybrid database, supports multiple models with model adapters, and uses, amongst others, methods such as truth discovery for conflict resolution and prioritization of data from chosen sources. Gaps in the data are addressed with external help.

Purpose

The work aims to answer three questions:

- How can cyber security modeling be made less manual and more automated to speed it up?
- How can (automated) modeling use multiple heterogeneous data sources?
- How can parts of heterogeneous data be prioritized for automatic modeling in such a way that knowledge about data sources is taken into account?

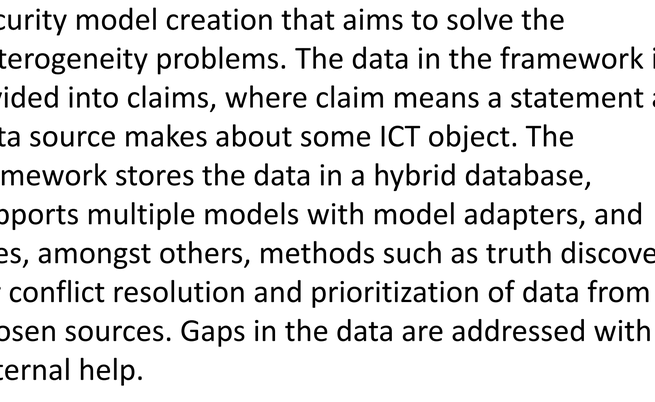

Figure 2. Small SCADA lab at KTH

Methods and Materials

- The work is based on design science [1].
- The problem has been defined in collaboration with stakeholders.
- New theories and artifacts (the framework) have been built using other theories and frameworks [2-9].
- The framework (Fig. 3) has been assessed and refined through a case study (described in Table 1) in a small SCADA lab (Fig. 2) at KTH.
- The theories have been implemented using Python, and a graph database ArangoDB (Fig. 4).
- The final IT infrastructure security threat assessment is done using securiCAD modeling tool.

Results

The results show that the quality of the output models depends on the quality of the input data. External help like an ontology is needed to improve data quality.

A probabilistic approach is preferable if there is doubt in the trustworthiness of the data sources. A truth discovery algorithm (like LCA [9]) can calculate the probability that a data source asserts correct claims, while prioritizing the claims from the data sources that have the most matching claims. The main benefit of using the truth discovery algorithm over naive versions is that it allows to use prior trust values and factor in uncertainty with negative claims. The negative claims ensure that claims from data sources that have been attributed the lowest trustworthiness values and that have the highest number of conflicts with other claims are removed .

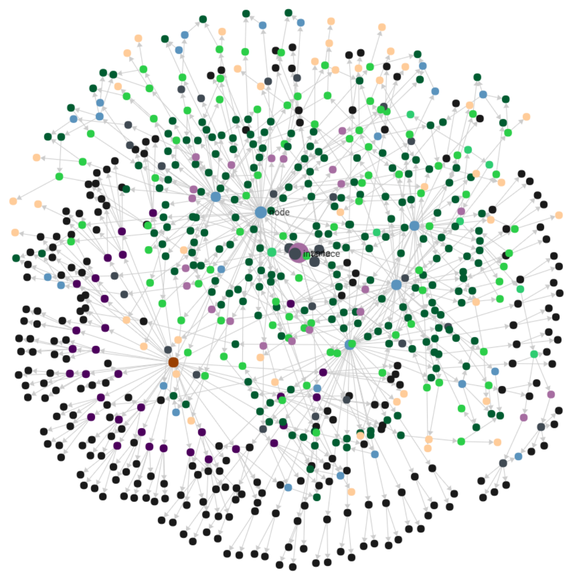

Figure 4. Data in graph database.

Table 1. Claims about IT architecture in the SCADA case study using Truth Discovery and Naïve analysis to resolve conflicts.

Name	Truth	Naive
Claims	1426	1426
Conflicting claims	2008	90
Non-conflicting claims	569	1336
Conflicting claims after conflicts solved	726	22
Saved to DB (without negative)	983	1358

Conclusions

Machine readable sources can be used as input for modeling, but heterogeneous data needs to be augmented and standardized, and conflicts need to be resolved. Data from different data sources can be prioritized using a truth discovery algorithm.

The framework was tested in a small SCADA lab with 3 different modeling tools to show the viability of the approach. During the case study two conflict resolution algorithms were compared – naïve (frequency based) and truth discovery (LCA) [9].

The tests show that the framework can be used to speed up IT architecture vulnerability assessments.

References

[1] R. H. von Alan, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” MIS Q., vol. 28, no. 1, pp. 75–105, 2004.
[2] E. Rahm and H. H. Do, “Data cleaning: Problems and current approaches,” IEEE Data Eng. Bull., vol. 23, no. 4, pp. 3–13, 2000.
[3] M. Farwick, B. Agreiter, R. Breu, S. Ryll, K. Voges, and I. Hanschke, “Requirements for Automated Enterprise Architecture Model Maintenance A Requirements Analysis Based On A Literature Review And An Exploratory Survey,” 13th Int. Conf. Enterp. Inf. Syst. (ICEIS), Beijing, pp. 325–337, 2011.
[4] M. Farwick, B. Agreiter, R. Breu, S. Ryll, K. Voges, and I. Hanschke, “Automation processes for enterprise architecture management,” Proc. - IEEE Int. Enterp. Distrib. Object Comput. Work. EDOC, pp. 340–349, 2011.
[5] M. Farwick, R. Breu, M. Hauder, S. Roth, and F. Matthes, “Enterprise Architecture Documentation: Empirical Analysis of Information Sources for Automation,” 2013 46th Hawaii Int. Conf. Syst. Sci., pp. 3868–3877, 2013.
[6] H. Holm, M. Buschle, R. Lagerström, and M. Ekstedt, “Automatic data collection for enterprise architecture models,” Softw. Syst. Model., pp. 825–841, 2012.
[7] M. Buschle, H. Holm, T. Sommestad, and M. Ekstedt, “A Tool for Automatic Enterprise Architecture Modeling,” CAISE11 Forum, p. 8, 2011.
[8] L. Berti-Equille and J. Borge-Holthoefer, Veracity of Data. Morgan & Claypool, 2016.
[9] J. Pasternack and D. Roth, “Latent Credibility Analysis,” in Proceedings of the 22nd international conference on World Wide Web, 2013.

SEGRID Security for smart Electricity GRIDs
www.segrid.eu info@segrid.eu

This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 607109