Margus Välja
2016/05/09

Data driven probabilistic cyber threat modeling

Architecture models are used in enterprise management for decision support. These decisions range from designing processes to planning for the appropriate supporting technology. Architecture models play also an important role in cyber security threat prediction and investment decisions. For mid-sized to large organizations model creation can be an enormous task when executed manually. Fortunately, there's a lot of data available from different sources within an enterprise that can be used for populating such models in a timely manner. The data are however almost always heterogeneous and usually only represent fragmented views of certain aspects. My research goal is to address this problem. The aim of my project is to devise a process to merge data for a cyber security threat prediction model in a way that the uncertainty in the data and the problems encountered during merging are visible in the end result.

The research project consists of the following phases.

1. Identify sources for data collection.
   There is a variety of potential data sources of enterprise data like network monitors and scanners, configuration management databases, project portfolio management tools and so on. The relevancy of each data source in modeling domain context should be studied.
2. Design the process for augmenting data using domain ontologies.
   The data from various data sources needs to be checked for consistency and the gaps in the data need to be addressed. Domain ontologies that describe different aspects of enterprise are fit for this purpose.
3. Design the probabilistic process from merging data that is able to reason over conflicts in unsupervised manner.
   A known statistical data analysis method like Bayesian Markov Chain Monte Carlo can be used to get a probabilistic overview of the problems in the merged data.
4. Implement the module as a prototype and test its accuracy in a corporate environment.
   The final step is the validation of the created module in a real environment.