



# ***Intrusion Detection via Machine Learning for SCADA System Protection*** paper commentary

Machine Learning Techniques

18<sup>th</sup> of December, 2016

Andrés Montoro Montarroso

[andres.montoro@alu.uclm.es](mailto:andres.montoro@alu.uclm.es)

Pedro-Manuel Gómez-Portillo López

[pedromanuel.gomezportillo@alu.uclm.es](mailto:pedromanuel.gomezportillo@alu.uclm.es)

## Link to the original paper

[http://ewic.bcs.org/upload/pdf/ewic\\_icscsr13\\_paper12.pdf](http://ewic.bcs.org/upload/pdf/ewic_icscsr13_paper12.pdf)

## Commentary

In today's world with such a growing industry, it is a constant necessity to be able to remotely control and supervise processes. *SCADA* systems, which stands for Supervisory Control And Data Acquisition, were born to give a proper solution to this problem.

Due to its nature, such systems are very susceptible to suffer from attacks of all kinds, and this characteristic make them very hard to protect. *IDS*, or Intrusion Detection Systems, were implemented in *SCADA* system for observing the events taking place on them. The principles of those systems is to detect as many attacks as possible, which drive them not to be very accurated and to generate a vast amount of data. But now that it has been possible to gather all this data appears the necessity to analyse them.

Notwithstanding the fact that several approaches have been implemented along the time to try to handle such data, seldom had these efforts performed ideally. For example, some years ago it was human analysts who evaluated the data and took decisions accordingly. Nevertheless, this difficult and time-consuming task do not make affordable for humans to perform it. Thus, it was needed a faster and more scalable approach.

*MLT*, or Machine Learning Techniques, have been reported to be able to solve this problem properly through several strategies. Two of the most popular ones are 1) *rule-based approach*, which tries to detect intrusions by recognizing inconsistent or anonalious behaviour on users and programs by means of comprehensive rules and 2) *support vector machines*, which tries to optimally classify the data into two clusters by mapping them into a space where it is linearly separable, even though it is very sensitive to noise.