21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015

Contribution ID: **14** Type: **oral presentation**

Intrusion Detection in Grid computing by Intelligent Analysis of Jobs Behavior –The LHC ALICE Case

Tuesday, 14 April 2015 14:45 (15 minutes)

Grid infrastructures allow users flexible on-demand usage of computing resources using an Internet connection. A remarkable example of a Grid in High Energy Physics (HEP) research is used by the ALICE experiment at European Organization for Nuclear Research CERN. Physicists can submit jobs used to process the huge amount of particle collision data produced by the Large Hadron Collider (LHC) at CERN. Grids allow the submission of user developed jobs (code and data). They also have interfaces to Internet, storage systems, experiment infrastructure and other networks. This creates an important security challenge. Even when Grid system administrators perform a careful security assessment of sites, worker nodes, storage elements and central services, an attacker could still take advantage of unknown vulnerabilities (zero day). This attacker could enter and escalate her access privileges to misuse the computational resources for unauthorized or even criminal purposes. She could even manipulate the data of the experiments.

Accordingly, Grid systems require an automatic tool to monitor and analyze the behavior of user jobs. This tool should analyze data generated by jobs such as log entries, traces, system calls, to determine if they run in a desired behavior or are performing any kind of attack on the system. The tool should react to the attack by sending alerts, logging information about relevant events and performing automatic defensive actions (for example stopping a suspicious process). This piece of software could
be classified as Grid Intrusion Detection Systems (Grid-IDS). Traditional IDS allow detection of attacks by fixed if-then rules based on signatures. It compares the input data with known predefined conditions from previous events. This strategy fails when a new type of intrusion is used, even with a slightly difference. Artificial intelligence algorithms have been suggested as a method to improve Intrusion Detection Systems. By the usage of a Machine Learning approach it is possible to train the IDS on generalizing among attacks even when they are completely new. Intelligent IDS can also analyze the huge amount of data generated in Grid logs and process traces to determine a misbehaving scenario (data mining). This Grid IDS has to be adapted to highly distributed scenarios, when collaboration among geographically separate sites is necessary and reliability on central services is not always an option.

Currently there is no framework that allows us to fulfill all the above requirements. We will design and build such framework. This framework should allow the monitoring and analysis of grid job behavior to detect attack attempts, even if new techniques or zero day vulnerabilities are utilized. This framework should also perform required countermeasures for its protection. In a first step, we plan to analyze the
behavior of the usual job execution in the ALICE experiment Grid. We will determine the most important metrics to characterize a "bad"behavior (an attack). Later we will collect data from the Grid logs using these metrics and will use this data to train a machine learning algorithm. The algorithm will allow us classification of jobs as in desired or undesired state depending on the data produced in their execution. We plan to implement the proposed framework as a software prototype that will be tested as a component of the ALICE Grid middelware.

Keywords – grid computing, distributed computing, distributed System security, artificial intelligence, data mining, Intrusion Detection Systems.

Primary author: GOMEZ RAMIREZ, Andres (Johann-Wolfgang-Goethe Univ. (DE))

Co-authors: LARA MARTINEZ, Camilo Ernesto (Johann-Wolfgang-Goethe Univ. (DE)); KEBSCHULL, Udo Wolfgang (Johann-Wolfgang-Goethe Univ. (DE))

Presenter: GOMEZ RAMIREZ, Andres (Johann-Wolfgang-Goethe Univ. (DE))

Session Classification: Track 4 Session

Track Classification: Track4: Middleware, software development and tools, experiment frameworks, tools for distributed computing