Contribution ID: **229**   Type: **Oral**

A scalable architecture for online anomaly detection of WLCG batch jobs

Thursday 21 January 2016 14:50 (25 minutes)

For data centres it is increasingly important to monitor the network usage, and learn from network usage patterns. Especially configuration issues or misbehaving jobs preventing a smooth operation need to be detected as early as possible. At the GridKa Tier 1 centre we therefore operate a tool for monitoring traffic data and characteristics of WLCG jobs and pilots locally on different worker nodes. On the one hand local information itself are not sufficient to detect anomalies for several reasons, e.g. the underlying job distribution on a single worker node might change or there might be a local misconfiguration. On the other hand a centralised anomaly detection approach does not scale regarding network communication as well as computational costs. We therefore propose a scalable architecture based on concepts of a super-peer network.

The contribution discusses different issues regarding the optimisation of computational costs, network overhead, and accuracy of anomaly detection. Based on simulations we will show the influence of different parameters, e.g. network size, location of computation, but also characteristics of WLCG batch jobs. The simulations are based on real batch job network traffic data that has been collected for several months.

Author: KUHN, Eileen (KIT - Karlsruhe Institute of Technology (DE))

Co-authors: PETZOLD, Andreas (KIT - Karlsruhe Institute of Technology (DE)); JUNG, Christopher; GIFFELS, Manuel (KIT - Karlsruhe Institute of Technology (DE)); FISCHER, Max (KIT - Karlsruhe Institute of Technology (DE))

Presenter: GIFFELS, Manuel (KIT - Karlsruhe Institute of Technology (DE))

Session Classification: Track 1

Track Classification: Computing Technology for Physics Research