Monitoring the network traffic of individual batch processes

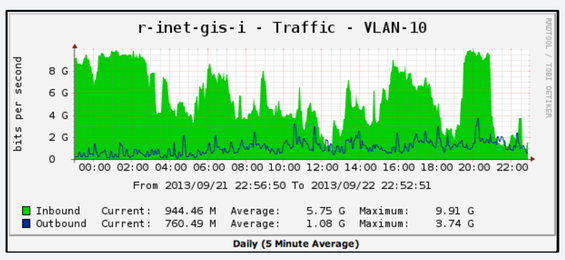
Eileen Kuehn, Karlsruhe Institute of Technology, Steinbuch Centre for Computing
eileen.kuehn@kit.edu

Motivation

- Federated Data Access introduces new workflows and data flows of batch jobs
- insight into internal and external network traffic usage requires additional monitoring tools
- condition: seemless integration into existing monitoring environments

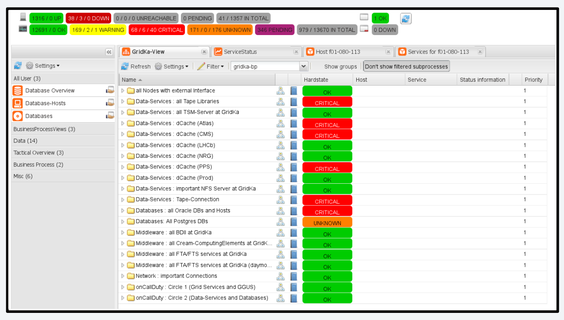
Background

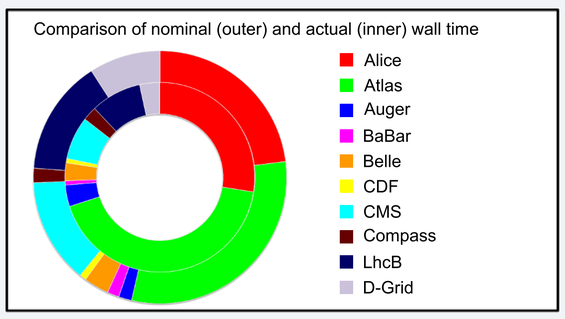
- different monitoring tools at GridKa
  - Cacti®
  - Icinga
  - Univa Grid Engine (UGE) monitoring
  - Ganglia



- Cacti® is a network graphing solution for the monitoring of network ressources
- actual traffic with focus on ports
- accumulated network traffic by rack

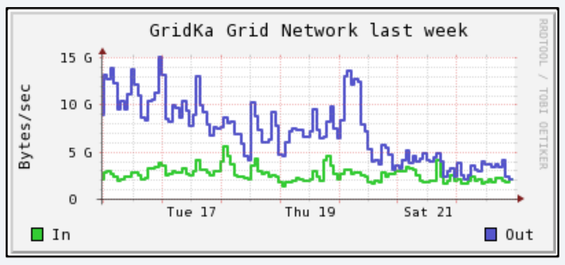
- Icinga watches all services including network ressources
- error notifications
- on-call alarms





- UGE displays statistics of running jobs
- data of finished jobs for accounting
- no information about incoming and outgoing traffic included

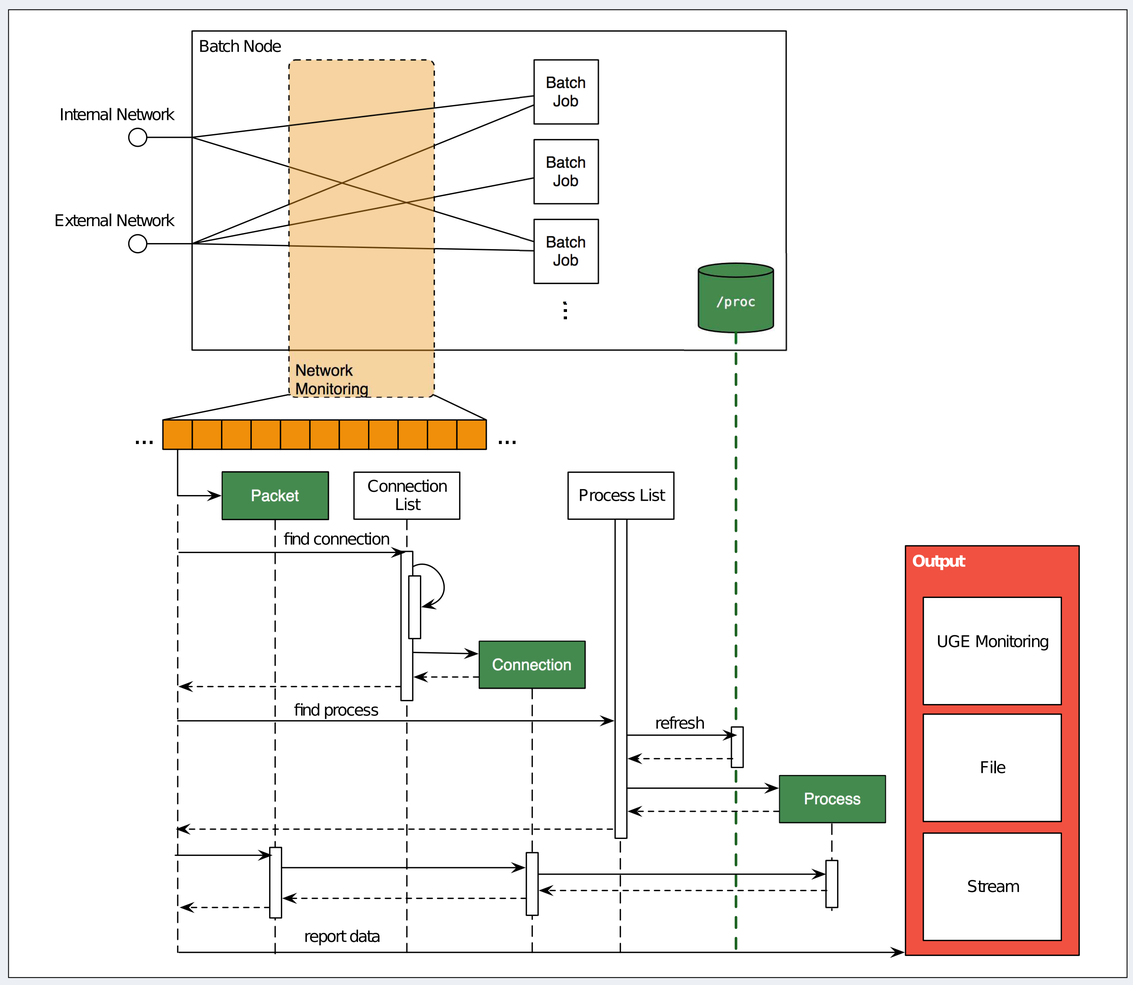
- Ganglia is used for performance metrics of nodes and services
- accumulated network traffic by node



Implementation

- currently no solutions concerning network traffic of batch jobs in monitoring environments
- new implementation and evaluations based on OpenSource tool NetHogs[1]
  - measuring bandwidth by process, not by protocol or subnet
  - relies on libpcap and ncurses
  - no need for special kernel modules

- monitoring of UDP/TCP packets
- splitting into internal/external network traffic
- group processes by their associated batch jobs
- configurable output to file/stream as csv or JSON
- final goal: active batch system plugin



Conclusion

- monitoring of internal/external network traffic of batch jobs
  - custom implementation based on NetHogs
  - implemented at GridKa Tier-1 center
- integration into existing monitoring environments and workflows by adding custom metrics to UGE
- output of measured data to file/stream for general purpose

1: http://nethogs.sourceforge.net