Holistic Performance Analysis of Large-Scale Distributed Systems

Francisco Neves and José Pereira
francisco.t.neves@inesctec.pt, jop@di.uminho.pt

Problem statement
The analysis of distributed systems is extremely challenging because:

• Distributed systems are built atop multiple heterogeneous third-party components.
• There is a lack of knowledge regarding the inter-component interactions.

Motivation

• The analysis of the behaviour of a system as a whole, including the interactions between its components, eases the effective reasoning about performance improvements.
• The existing tools are not suitable to address this issue as they focus on single components and thus inter-component interactions remain hidden.

Research Direction

• **Develop a black-box approach for discovering and monitoring the behaviour of a system.**
• **Black-box** stands for operating without instrumenting the application’s components.

The resulting monitoring system:

• Leverages the data captured by existing OS-level tracing tools.
• Coherently combines the collected data according to the system’s runtime behaviour.
• Provides a unified and extensible API for querying the collected data.

Representation of the architecture and interactions between components of a distributed system. The graph of processes illustrates the interactions of running processes.

With this monitoring system, we are able to answer questions as follows:

• What is the energy consumption of the complete stack used by an application?
• What services should be affected by a shutdown of the air conditioning system in a given datacenter hall?