

## Software Fault Diagnosis

Software fault diagnosis is concerned with determining the **root cause** and possible **repair actions** for an observed **software failure**.

Although the term **software fault** is typically used to mean **software defect**, software failures are not always caused by software defects. Other **likely causes**, which cannot be simply removed by software developers, include:

- Missing runtime resources and problems in the environment
- Misconfiguration of interacting components
- Misuse of software against specifications
- Missing understanding of operational issues

The **observable feedback** produced by failing software is often inadequate to quickly determine the fault. This problem is aggravated by operators' **insufficient knowledge** of the software's inner structure, which remains **hidden and irrelevant** during normal operation.

The knowledge required to successfully diagnose software faults in widely deployed software is often available online because the same faults become activated across installations. Unfortunately, usual representations of such knowledge make it inaccessible to **automated troubleshooting agents**.

## Grid Middleware

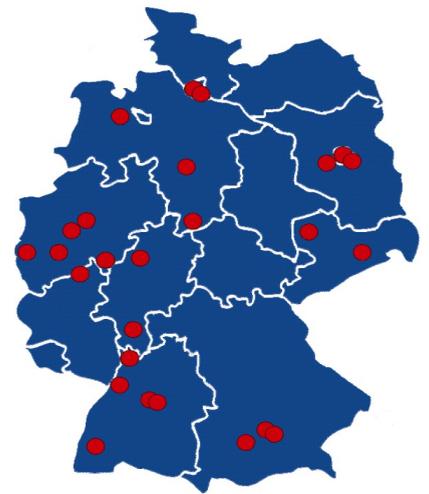
**Grid middleware** is a category of software used to provide access to distributed high-performance computing and data storage. In the research project **WISENT** (<http://wisent.d-grid.de>), Grid middleware is deployed to support the scientific collaboration of the energy meteorology community, concerned with research such as availability forecasting of renewable energy sources.

Basic user interactions with Grid middleware include

- discovering available resources
- submitting computational jobs
- requesting data transfers

Today, many different implementations of Grid middleware exist and must be supported and integrated to provide service to disparate user communities. Examples include Globus, UNICORE and gLite, along with workload management systems such as PBS/TORQUE, Condor, SGE, LSF and others.

The relative **novelty**, **distributed character**, **heterogeneity** and **intended users** of Grid middleware make software fault diagnosis an especially challenging and relevant task in this context.



Computing centers participating in the German Grid  
Source: D-Grid Initiative

## Bayesian Networks

**Bayesian networks** originate from research on probability theory and Artificial Intelligence. They are a powerful notation, which can be used for representing causal relationships among a set of uncertain variables. Each variable represents a state of the world in which reasoning is performed, which may be unknown to the reasoner.

To aid diagnosis, probabilities that describe relationships between nodes in a Bayesian network may be interpreted as **degrees of belief** of an expert faced with an inference task rather than long-run frequencies in a statistical sense.

Bayesian networks have been successfully applied for diagnostic tasks in medicine and hardware troubleshooting. In project WISENT we explore their possible applications to the problem of software fault diagnosis.

One idea is a semi-interactive user support system which relies on a Bayesian network to accumulate knowledge about faults and observable symptoms and to guide observations and repair actions (cf. Heckerman 1995).

