In Search of Real Data on Faults, Errors and Failures

Moderator:
Miroslaw Malek (Humboldt-Universität zu Berlin, Germany)

Panellists:
Domenico Controneo (Università di Napoli, Federico II, Italy)
Zbigniew Kalbarczyk (University of Illinois Urbana-Champaign, USA)
Dave Penkler (Hewlett-Packard, France)
Manfred Reitenspiess (Fujitsu Siemens Computers, Germany)

In order to make a relevant contribution to industrial practice and have an impact on the future systems and services, it is essential that the research community has access to real data to be able to test effectiveness and verify correctness of proposed techniques for enhanced availability (the term “real data” refers to the field data collected at the customers’ sites, not just at the lab where usually experts assume various scenarios without proper attention to operator mistakes, environment and customer’s maintenance procedures). To date the community had rather sporadic opportunities to access of the field data and have developed the body of knowledge based frequently on wrong assumptions, hypothetical failure models and simplistic distributions. At the core of the problem is that the failure data are classified due to the competition and the fact that almost always it is attached to specific customers and its bulk may be enormous. With thousands of measurement points and up to about 1200 parameters that can be measured on computer and communication systems, the amount of data may reach from several Gbytes to over a hundred Gbytes per day. The key challenge is how to filter out real data and code it such that it can be accessed by the research community while at the same time the bulk of data is significantly reduced by focusing strictly on faults, errors and failures and their root causes.

To change this state of affairs, the panel will attempt to give pointers to the sources of real data, investigate the ways of collecting the data and making it accessible by the research community. The panel includes academic and industrial experts.