

## **Availability in large distributed systems**

*Bruno Gaujal*

Abstract:

In the age of cloud, Grid, P2P, and volunteer distributed computing, large-scale systems with tens of thousands of unreliable hosts are increasingly common. Invariably, these systems are composed of heterogeneous hosts whose individual availability often exhibit different statistical properties (for example stationary versus non-stationary behavior) and fit different models (for example exponential, Weibull, or Pareto probability distributions). In this talk, I will describe an effective method for discovering subsets of hosts whose availability have similar statistical properties and can be modelled with similar probability distributions. These methods and models are then used for the design of asymptotically optimal scheduling algorithms across large systems where host availability is uncertain.