Chapter 1
Towards a Holistic Approach to Fault Management: Wheels Within a Wheel

Moises Goldszmidt
Microsoft Corporation, USA

Miroslaw Malek
Humboldt-Universität zu Berlin, Germany

Simin Nadjm-Tehrani
Linköping University, Sweden

Priya Narasimhan
Carnegie Mellon University, USA

Felix Salfner
Humboldt-Universität zu Berlin, Germany

Paul A. S. Ward
University of Waterloo, Canada

John Wilkes
Google Inc., USA

ABSTRACT

Systems with high dependability requirements are increasingly relying on complex on-line fault management systems. Such fault management systems involve a combination of multiple steps – monitoring, data analysis, planning, and execution – that are typically independently developed and optimized. We argue that it is inefficient and ineffective to improve any particular fault management step without taking into account its interactions and dependencies with the rest of the steps. Through six real-life examples, we demonstrate this inefficiency and how it results in systems that either under-perform or are over-budget. We propose a holistic approach to fault management that is aware of all relevant aspects, and explicitly considers the couplings between the different fault management steps. We believe it will produce systems that will better meet cost, performance, and dependability objectives.

INTRODUCTION

Large, complex systems frequently experience faults, and those faults need to be handled to limit the damage they cause. Fault management is the set of processes used to ensure dependability of the service, i.e., uninterrupted, reliable, secure and correct operation, at reasonable cost. An important sub-goal in modern fault management approaches is to automate as much as possible to
reduce human intervention and administration, which is expensive, error-prone and may even be infeasible in large systems. In turn, that requires a clear statement of the objectives and processes used to achieve the desired outcomes.

Fault management systems typically include many steps, which are often carried out sequentially, as shown in Figure 1: system monitoring, analysis of monitored data, planning of recovery strategies, and execution of mitigation actions. Although this taxonomy is a convenient compartmentalization, it is usually ineffective and inefficient to optimize a particular fault management step in isolation, without taking into account its interaction with the rest of the steps and how these interactions affect the overall objectives. That is, rather than asking “how should we improve a particular step?” a better question is “how should we configure the steps to maximize the overall benefit?”

For example, if the planning step offers only three possible recovery actions – say, to reboot a machine, reimage a machine, or call the operator – it is unnecessary for the analysis step to do anything other than to map the outcome of the monitoring step to one of these three actions. Any further effort in the analysis step is irrelevant as it has no bearing on the overall fault management outcome; indeed, such further effort might complicate the overall system (e.g., introduce bugs), waste cycles at runtime (which may impact availability), and waste developers’ time on meaningless tasks.

We propose a holistic approach to the problem of deciding how much effort to invest where by addressing all four steps (monitoring, analysis, planning and execution) and their influences on each other, keeping in mind the main objectives – cost minimization and high availability. Our approach allows local optimization within the operational envelope of each fault management step, but links the global adaptation of these local optimizations to the (global) business goals of the system, resulting in a highly effective and coordinated combination of the four fault management steps. Therefore, we avoid local optimizations that do not help achieve the overall goal.

In support of our arguments, we present six real-life examples that explore the pitfalls of merely-local optimizations. The price of not having a proper focus on key objectives and ignoring the holistic approach is high and can no longer be neglected. This paper is a call to arms, to energize the community and give the impetus needed to overcome pitfalls of local optimization,