

Fault and Performance Management for the Access Grid

Cynthia S. Hood
Illinois Institute of Technology
hood@iit.edu

1. Introduction

Our national computing infrastructure is comprised of a set of diverse computing and network resources with varying levels of performance and reliability. Ideally, we would like to utilize these resources efficiently while providing adequate service to the users. To accomplish this, effective fault and performance management techniques must be developed.

The distinction between fault and performance management is fuzzy. In our research we broadly define a problem as anything that has a negative impact on performance. Our goal is to automatically detect subtle clues that problems may be occurring and make appropriate corrections to avoid significant service degradation.

We are currently studying this problem in the context of the Access Grid (AG). The AG provides the opportunity to explore a live, large-scale system. In this paper we will describe the ideas being studied as well as the approach.

2. Fault and Performance Management Issues

Fault management has traditionally been defined as the detection, diagnosis, and correction of a fault or problem. Performance management is the detection and correction of a performance problem. Typically the system is monitored to enable automatic detection. The implication (and in fact practice) is that there is a progression of events; first a problem is detected, then the problem is diagnosed, and then corrective actions may be taken. The issue with this approach is that it implies that a problem must be diagnosed before corrective actions may be taken.

Diagnosing a problem is very difficult because of the complexity of the systems and the constant change. The systems are complex functionally as a whole, and additional complexity is introduced through the interactions of heterogeneous components. The change comes about both internally through hardware and software updates and new components, and externally through environment. The behavior of the system is changing at different rates due to load, traffic mix and configuration. Typically diagnosis involves the correlation of information from a variety of sources. Issues related to the correlation problem include:

Patterns/Models – most of the approaches to correlation involve matching gathered information with known patterns or models of faults. However, it is *infeasible* to have a complete set of models, and the models *change over time* and are dependent on several things including configuration, traffic mix, and load.

Scope – at what level should information be correlated (depth), and how much information should be correlated (breadth)?

Synchronization – information is received at different times, and information from some sources may be incomplete or completely missing.

Security – in a distributed environment, systems are typically unwilling to share fault/performance monitoring data.

Due to the complexity of these issues, little progress has been made in the event correlation arena. For this reason, we propose a fundamentally different approach to the fault management problem.

3. Our Approach to Fault and Performance Management

Our approach focuses on the decision points within the system rather than focusing on the faults. Instead of trying to figure out exactly what is going on in the network, we focus on making good decisions given the information available. This significantly reduces the complexity of the problem since it involves focusing on the decision space rather than the fault space. This is a very reasonable approach, particularly given that corrective actions are ultimately constrained by the types of decisions and corrective actions available. Further, techniques focusing on making good decisions with limited information are of critical importance given that global information may not be available due to network congestion or failure.

The two main pieces to this research are: (1) determining how simple agents can make good decisions at the decision points and (2) understanding how to coordinate the decisions made by the agents at a strictly local level to achieve good global results. Since decision points exist at multiple levels in any system, our approach processes information and acts at several different levels of abstraction. The goal is to fix the problem at the appropriate level of abstraction, and to make local decisions that do not degrade global performance.

Statistical methods fit well with this approach. At each decision point, the agents will extract the information pertinent to making a good decision. Our thesis is that this will amount to differentiating between local problems that can be fixed at the decision point on a local level and symptoms of more global problems. By analyzing the quality of decisions made with local information vs. global information correlated from multiple sources, we will determine how much correlation is necessary and what information should be correlated. A key benefit of this approach is that the amount of correlation

necessary will depend on the complexity of the decision rather than the complexity of the fault.

4. The AG as a Research Tool

For many reasons, the AG provides a great opportunity for this type of research. One of major obstacles to fault and performance management research is the availability of data. For commercial systems and networks, fault and performance data is considered highly proprietary and typically not shared (many times not even within a company). The AG can provide a significant source of “real” data. In addition, the data can be collected from many different levels, providing several different viewpoints. By collecting performance data from vic and rat, we can label the data collected at other levels. This is possible because there are well-defined performance parameters for these types of applications (i.e. streamed audio and video), unlike most applications. We expect to make the data collected publicly available. The data will be cleansed to address any security concerns.

The use of public domain software allows us to make modifications both to collect the data and to eventually implement solutions. We can get into the code and identify decision points that are not visible to the user. As the project progresses, we expect to run experiments on the AG. Given the research nature of the AG, we anticipate cooperation and are looking forward to collaborating with others.

5. Current Status

We are currently working in three different directions to forward this research. The first direction is data collection. We are interested in collecting any of the information that is available and are working on putting an infrastructure in place. Statistics are collected by vic and rat, but we need to make modifications to store the statistics (they are currently displayed on the screen). We are working to collect various types of end-to-end and network information. We are hoping to be able to collect WAN statistics from Abilene. Any help in this area would be greatly appreciated –AG node specific information will enrich the research.

The second direction is identification of the decision points. We are working to identify the decision points within vic and rat, and also within the network. Once significant progress is made on identification, modeling will begin. We are beginning with decision points that are externally visible (i.e. those that an operator could control).

The third direction is simulation of the AG. We are working on an OPNET simulation. This is a critical piece for understanding the interactions of the decisions, especially the decisions made in the network.

Acknowledgements

Thanks to Bill Gropp for pointing me to the AG and supporting this research through the Argonne Summer Faculty program. This work is also partially supported through NSF 9984811.