# IBM Research Report

## Autonomic Computing Features for Large-Scale Server Management and Control

**R.K. Sahoo, I. Rish, A. J. Oliner,**
**M. Gupta, J.E. Moreira, S. Ma**
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598-0218

**R. Vilalta**
Department of Computer Science
University of Houston, 4800 Calhoun Rd, Houston, TX 77204

**A. Sivasubramaniam**
Department of Computer Science and Engineering
Penn. State University, College Park, PA

# Autonomic Computing Features for Large-scale Server Management and Control

R.K. Sahoo, I. Rish, A. J. Oliner *, M. Gupta, J.E. Moreira, S. Ma
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598-0218
{rsahoo,rish,ajoliner,mgupta,jmoreira,sma}@us.ibm.com

R. Vilalta
Department of Computer Science
University of Houston, 4800 Calhoun Rd, Houston, TX 77204
vilalta@cs.uh.edu

A. Sivasubramaniam
Department of Computer Science and Engineering
Penn. State University, College Park, PA
anand@cse.psu.edu

## Abstract

A computer system would satisfy the requirements of "*autonomic computing*", if the system can configure and reconfigure itself by knowing the operating environments, protect and heal itself from various failures or malfunctions. In order to know the environments and detect failure, an autonomic system needs the capability of acquiring the information through self-monitoring.Once the sequence of events leading to a series of disasters are figured out, it is required to predict and control the system management process through a number of automated learning and proactive actions.

In this paper, we address the cluster system RAS (Reliability Availability and Serviceability) by analyzing the realistic system event log history, collected from a 250 node large-scale cluster. Based on the analysis of these events through a number of machine-learning and artificial intelligence techniques, we have established the scope of time-series methods, rule-based classification techniques and Bayesian network algorithims for overall self-management and control. While the time-series methods can be used effectively for predicting system performance parameters, the rule-based classification algorithms effectively implemented to predict future critical events up to 70% accuracy. Bayesian network based algorithms can be used for root-cause analysis through adaptive probing and establishing probe managers.

We also cover some of the ongoing efforts to provide an online prediction and control mechanism through a hybrid model combining the selected artificial intelligence and machine learning techniques including active probing and triggers.

## 1   Introduction

Large-scale clustered systems like Blue Gene [1] can have up to hundred to thousands of system units. Managing the status and usage of these systems may consist of sets of very complex and labor-intensive processes. Design-

---

1

ing an effective autonomic system management for large-scale clustered system would ensure the proper usage of the systems, which would also translate into significant cost saving. Autonomic system management may include many features like self-configuration, self-optimization, self-healing and self-protection. These proactive prediction and probing capabilities will provide the system management components with the pertinent information such that self-configuration, self-healing and self-optimization are possible for critical system resources. Our technical treatment of autonomic monitoring is based on our experiences on large scaled clusters for scientific and technical computing (Blue Gene/Light). Specifically, we discuss the applicability of a number of time-series, rule-based classification algorithms for prediction and root-causal analysis through Bayesian network techniques applicable for large-scale systems.

## 1.1 Cluster System predictive failure analysis

The malfunctioning of any node in a system can be diagnosed through a heartbeat monitoring process. Hence heart beating is one of the first step towards achieving self-management for large-scale clusters. However, heart beating alone cannot provide any detailed information or the path of the error or fault propagation within a large cluster. Apart from heart-beat monitoring significant amount of analysis and control mechanisms are required to automate the process of system management for large-scale servers.

For large-scale clusters this sort of problems become easily complex due to the simultaneous reporting of a wide range of hardware and software components. In order to address such a complex problem a machine learning algorithm based approach is carried out to make the system prediction more intelligent, and to reduce failures.

## 2 Event log collection for Data Analysis

Based on sysem activity and unusual event logs collected from a 250 node based cluster, we have established a list of primary and derived variables for use within the scope of the AI and machine-learning algorithms. While the primary variables provide the default information about the machine related parameters, the derived variables try to extract some of the hidden features within the parameters by processing and establishing the inter-parameter relationships. Some of the primary variables are *Time-Stamp*, *Severity*, *Node ID* etc., *Delay* inter-arrival time between two events, while *Global inter-arrival time of the events* (either within the same node ID or same event type), are considered as derived variables. More details of the data collection and filtering the data to establish the distinct series of events are discussed elsewhere [5, 6].

## 3 AI Algorithms and Cluster System RAS

A number of time-series and belief-network algorithms are available in the literature applicable to RAS event analysis. We confine our analysis to three types of algorithms: (1) Time-series algorithms, 2) Rule-based classification algorithms and (3) Bayesian network algorithms. The algorithms are chosen based analysis response of the cluster based collected data and the applicability of the algorithms.

## 3.1 Time-series algorithms

Linear time-series models [3] have been successfully used for forecasting and prediction in various fields. We use time-series models to predict system parameters like percentages of system utilization (*%sys*), idle time (*%idle*), and network I/O (*%IO*). For initial calculations, we assume the events to be distributed at equal time intervals, so that the corresponding system scalar functionalities can be easily used as input parameters for time-series models.

Based on single node based analysis, it has been established that:

- Overall, the *LAST* model does better than other time-series models. This is mostly due to the small changes associated with the performance parameters, compared to the way event logs change with time.

- For continuous data like *%sys*, the mean error decreases monotonically as the size of the training

dataset increases [6].

## 3.2 Rule-based Classification Algorithms

One approach to predict computer system events follows an association rule-based strategy. The idea is to first identify those events of a critical nature that are important to predict; these events belong to a category that may cause serious performance damage. These critical events become our target for prediction. To proceed, eventlogs are analyzed by looking for patterns, or sets of events, that frequently precede target events. These patterns can serve as precursors: an alarm system can be implemented to raise a flag whenever any of these precursors is detected. Before the construction of alarm rules, patterns are validated to ensure they genuinely precede target events; patterns must not be part of background noise to be used for prediction. The end result is a set of temporal association rules that can anticipate the occurrence of target events. Experimental results using a large-scale cluster of nodes have shown to predict rare events up to 70% accurate based on rule-based algorithms [6, 7].

## 3.3 Bayesian network techniques and root cause analysis

In this section, we describe our progress to establish the root cause of events through learning probabilistic dependency models, such as Bayesian networks, from event data. A Bayesian network model [4] describes domain variables (such as event occurrence, event severity level, etc.) and probabilistic dependencies among specified by conditional probability distributions. Bayesian networks provide a compact representation of multi-variate joint distributions and support efficient algorithms for inference tasks such as prediction and diagnosis.

As a part of our initial analysis we attempted to reconstruct the dependencies between the primary and derived variables for the cluster event data. First, we focused only on variables describing the events coming from a single node (called herein the 'single-node' analysis). Then we considered all variables describing the whole cluster ('cluster analysis'). The results were obtained using B-Course [2], an interactive web-based tool that allows to learn Bayesian network models from data and to perform inferences based on observations.
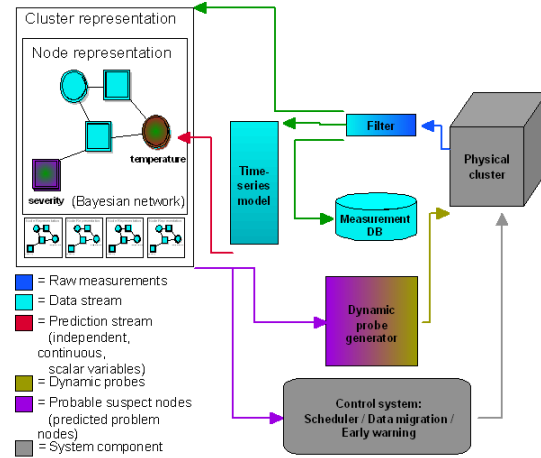


Figure 1: Hybrid Prediction Model

# 4 Hybrid Prediction and Proactive Control Model

By combining the various methods of analysis, we have designed a hybrid prediction system for large clusters (Figure 1). The prediction system begins in a tabular state. The model is not built based on any assumptions about the conditions under which errors will occur or the behavior of the independent variables of the cluster. Over time, the model learns the patterns and would be able to flag the nodes, when it believes to have a high probability of failure or occurrence of critical events.

The system reports two types of information: (1) Events, (2) States. The events are generally in the form of event logs. While the states constitute other health related signals such as : temperature, CPU utilization etc.

For each node in the cluster, where a node is the smallest set of components for which the system differentiates, we maintain a static Bayesian network, called the "*Node Representation*" (*NR*). A set of node representation is defined as a "*Cluster Representation*" (*CR*). The error or event stream output by data filter is used as the training data for the node, in which the event occurred. We only maintain an (*NR*) for those nodes in which at least one error has occurred, since the system will often be error-free thus maintaining an up-to-date adaptive node related information.

The state information output from the filtered data, can be further categorized to isolate these variables into either continuous, independent valued variables. This stream of values can be used as input to time-series tools, which keeps a mathematical model of the behavior of each of these variables for each node. Thus the time-series based model can, not only continue to refine its model as new filtered data sets arrive, but also can make predictions about the future state of the variables. These predictions ( in terms of false positives or false negatives) which can be used by the associated *NR*. The *NR* given a predicted future state, can associate the events through Bayesian network to represent probabilistic cause of any particular type of events.

The *NR*s and dynamic *CR*s can help to setup effective probing mechanism by notifying the control system appropriately. Moreover, any kind of job scheduling and data migration choices can be carried out depending on the existing policies. As a result of the actions of the control systems, and probing, the database would gather information about the node health and in a sense providing more information about the system.

## 5 Summary

This paper describes the overall summary of an ongoing project to design and develop a proactive RAS system analysis, prediction and control mechanism for large scale clusters. The goal of the project is to define and implement autonomic computing features for large-scale system reliability availability and serviceability (RAS) for proactive system management and control. We concentrate our work addressing the following aspects of system RAS through realistic data analysis, experimentation and use of a number of artificial intelligence and statistical algorithms.

Within a broad category of system management and control, we concentrate our work focussing on (1) Effective RAS data retrieval and usage, (2) Prediction of critical events through rule-based classification algorithms, (3) Root cause analysis through Bayesian network techniques. As an ongoing effort, we are also experimenting and carrying out algorithmic developments to make the overall system management and control to be an automated process through a hybrid model.

The hybrid proactive system management and control model aims at establishing a complete end-to-end automated system management and control including establishing adaptive probes through "Probe Manager" and defining the policies based on the historic data.

Our future work includes alternative ways to formulate the whole hybrid-model to make the system management process more effective and online. It also includes establishing a well defined Bayesian network based root cause analysis component within the "Cluster Representation" (CR) and "Node Representation" (NR).

## References

[1] N. R. Adiga, et al. An Overview of the BlueGene/L Supercomputer. *Supercomputing (SC2002)*, 2002.

[2] A web-based Data-analysis tool for Bayesian Modeling. *http://b-course.cs.helsinki.fi.*

[3] P. J. Brockwell, R. Davis. Introduction to Time-Series and Forecasting. *Springer-Verlag*, 2002.

[4] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, San Mateo, California, 1988.

[5] R. K. Sahoo, M. Bae, R. Vilalta, J. Moreira, S. Ma and M. Gupta. Providing Persistent and Consistent Resources through Event Log Analysis and Predictions for Large-scale Computing Systems. *SHAMAN, Workshop, ICS'2002*, New York, June, 2002.

[6] R. K. Sahoo, I. Rish, A. J. Oliner, M. Gupta, J. Moreira, S. Ma, R. Vilalta and A. Sivasubramaniam. Critical Event Prediction for Proactive Management in Large-scale Computer Clusters. *Submitted to KDD-2003*.

[7] R. Vilalta, S. Ma. Predicting Rare Events in Temporal Domains. *Proceedings IEEE Conf. on Data Mining (ICDM'02), Japan.*