CS595: Proposal of Course Project. Due Oct 8, 2008

[Project Title]

Student Name (Last, First)

Abstract

(1 paragraph)

Table of Contents

1. Introduction 2
2. Subject Details 2
   2.1 Failure logs Characteristics 2
   2.2 Models and Algorithms to analyze failures…… 2
3. Comparison or Examples 2
4. Summary (optional) .....**Error! Bookmark not defined.**
5. References (optional) 2

1.Introduction

In the next few years production systems are expected to contain tens to hundreds of thousands of computing nodes and thousands of I/O nodes [21]. Such a scale, combined with the ever-growing system complexity, is introducing a key challenge on fault management in high performance computing (HPC). Despite great efforts on the design of ultra-reliable components, the increase of system size and complexity has outpaced the improvement of component reliability. Recent studies have pointed out that the mean-time-between-failure (MTBF) of teraflop and soon-to-be-deployed petaflop machines are only on the order of 10 - 100 hours.

To better fault tolerant these failures, we need first understand the character of these failures. These project aims to using scientific computing method to analyze the failure logs and try to find the some rules or models behind the failures.

2.Subject Details

2.1 Failure logs Characteristics

From the historical failure logs, we can analyze its characteristics, like its event time, its entry data, its occurred location.

2.2 Models and Algorithms to analyze failures

Exponent distribution:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & , \ x \geq 0, \\ 0 & , \ x < 0. \end{cases}$$

Weibull distribution:

$$f(x;k,\lambda) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Gamma distribution:

$$g(x;\alpha,\beta) = x^{\alpha-1}\frac{\beta^{\alpha}\, e^{-\beta\, x}}{\Gamma(\alpha)} \text{ for } x > 0.$$

To use these 3 distribution method, we will try to find a fit distribution for the failure logs. Then this model can be used to predict further failures occur time.

3.Comparison

We will compare these 3 method, to better understand the historical failure logs. Also we will compare these analyze method to existence academic method and try to give some advantage and disadvantage of this distribution method.

5.References

Jiexing. Gu, Z. Zheng, Z. Lan, J. White, E. Hocks, B.H. Park, “Dynamic Meta Learning for Failure Prediction in Large Scale Systems: A Case Study”, Proc. of International Conference on Parallel Processing (ICPP), 2008.

R.K. Sahoo, A.J. Oliner et al., “Critical event prediction for proactive management in large-scale computer clusters”, Proc. of KDD, 2003, pp. 426-435.