Reliability Engineering and System Safety 104 (2012) 96–105

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

An automated method for estimating reliability of grid systems using Bayesian networks

Ozge Doguc*, Jose Emmanuel Ramirez-Marquez

School of Systems & Enterprises, Stevens Institute of Technology, Hoboken, New Jersey, United States

A R T I C L E I N F O

Article history:
Received 16 February 2010
Received in revised form
11 March 2012
Accepted 14 March 2012
Available online 18 April 2012

Keywords:
Grid systems
Bayesian networks
Reliability
Grid service
Minimum resource spanning tree

A B S T R A C T

Grid computing has become relevant due to its applications to large-scale resource sharing, wide-area information transfer, and multi-institutional collaborating. In general, in grid computing a service requests the use of a set of resources, available in a grid, to complete certain tasks. Although analysis tools and techniques for these types of systems have been studied, grid reliability analysis is generally computation-intensive to obtain due to the complexity of the system. Moreover, conventional reliability models have some common assumptions that cannot be applied to the grid systems. Therefore, new analytical methods are needed for effective and accurate assessment of grid reliability. This study presents a new method for estimating grid service reliability, which does not require prior knowledge about the grid system structure unlike the previous studies. Moreover, the proposed method does not rely on any assumptions about the link and node failure rates. This approach is based on a data-mining algorithm, the *K*2, to discover the grid system structure from raw historical system data, that allows to find minimum resource spanning trees (MRST) within the grid then, uses Bayesian networks (BN) to model the MRST and estimate grid service reliability.

1. Introduction

Grid computing has become relevant due to its applications to large-scale resource sharing, wide-area information transfer, and multi-institutional collaborating. In general, in grid computing services request a set of resources, available in a grid, to complete certain tasks. Many experts believe that the grid technologies will offer a chance to extend the benefits of the Internet [1]. However, it is difficult to analyze the grid reliability due to its highly heterogeneous and distributed characteristics. Because the grid systems involve cross-organizational sharing, they support existing distributed computing technologies. As an example, enterprise-level distributed computing systems can use the grid technologies to achieve resource sharing across its different institutions. Although, several development tools and techniques for the grid systems have been studied, estimating grid reliability is not straightforward due to the size and complexity of the grid [2]. Therefore, new analytical methods are needed to evaluate the grid reliability.

Over the past several years, research and development efforts have focused on the challenges that arise when large grid organizations [1–4] are built. As a recent topic, there are a few studies on estimating grid system reliability in the literature [5–8]. In these studies, the grid system reliability is estimated by focusing on the reliabilities of services provided in the grid system. For this purpose, the grid system components that are involved in a grid service are classified into spanning trees, and each tree is studied separately. However, these studies mainly focus on understanding grid system structures rather than estimating the actual system reliability. Thus for simplification purposes, they make certain assumptions on component failure rates, such as satisfying a probabilistic distribution [7].

For reliability estimation, Bayesian networks (BN) have been proposed as an efficient method [9–12]. BN provide significant advantages over traditional frameworks for the systems engineers, mainly because they are easy to interpret and they can be used in interaction with domain experts in the reliability field [13]. Using the BN structure and the probabilistic values, the system reliability can be estimated with the help of Bayes rule [12]. There are several recent studies for reliability estimation using BN [9,11,14–16], which require specialized networks that are designed for a specific system. That is, the BN to be used for analyzing system reliability should be known beforehand (i.e. the BN can be built by an expert who has "adequate" knowledge about the system under consideration). However, human intervention is

Abbreviations: BN, Bayesian Network; RST, Resource spanning tree; MRST, Minimum resource spanning tree; CPT, Conditional probability table; K2, named after Kutat 2; QoS, Quality of service; RM, Resource manager; RN, Root node

* Corresponding author. Tel.: +1 201 920 4332; fax: +1 201 920 4641.
E-mail addresses: ozgedoguc@hotmail.com, odoguc@stevens.edu (O. Doguc).

0951-8320/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2012.03.016

Nomenclature

G_i	Component i in the grid system
S_i	Service i in the grid system
R_i	Resource i in the grid system
u	Maximum number of parents in the BN
T	Historical dataset
P_i	Set of parents of node i in the BN
t_i	Time that observation i was done on the grid system
f	Scoring function for the $K2$ algorithm
m	Number of nodes in the parent set
π_i	Set of parents for node i in the BN
rel	Doguc and Ramirez-Marquez's method for estimating reliability
combine	Dai and Wang's method for combining reliabilities

always open to unintentional mistakes that could cause discrepancies in the results [17].

To address these issues, this paper introduces a methodology for estimating grid system reliability by combining techniques such as BN construction from raw component and system data, association rule mining and evaluation of conditional probabilities. Based on the extensive literature review, this is the first study that incorporates these methods for estimating grid system reliability. With the increasing popularity of computer environments in systems engineering, grid systems have been widely used in various system-related applications. Understanding the grid system structure and the component relationships is essential for systems engineers for optimal resource allocation and improving the system reliability. This study provides a methodology for automated discovery of component relationships and estimation of reliability of grid services to help the systems engineers.

The methodology suggested in this paper automates the process of spanning tree discovery and BN construction by using the $K2$ algorithm (a commonly used association rule mining algorithm) that identifies the associations among the grid system components by using a predefined scoring function and a heuristic. According to the proposed method, once the BN is efficiently and accurately constructed, reliabilities of grid services are estimated with the help of Bayes rule. Unlike previous studies, the methodology proposed in this paper does not rely on any assumptions about the component failure rates in grid systems. Moreover, the proposed method does not require prior knowledge about the grid system structure.

2. Background information

This section provides background information about the grid systems, BN and the $K2$ algorithm. Earlier studies on estimating grid system reliability are also discussed in this section.

2.1. Grid systems

To represent distributed computing infrastructures for advanced science and engineering, the term "grid" was first used in the 90s [3]. The grid concept was first developed to enable resource sharing within geographically diverse scientific organizations. The main problem that lies under the concept of grid systems is coordinated resource sharing and problem solving in dynamic and multi-institutional organizations [1]. Different than typical distributed systems, the computational grid systems require large-scale sharing of resources on different types of components. A service request in a grid system involves a set of nodes and links, through which the service can be provided. In a grid system, the Resource Managers (RM) control and share resources, while the Root Nodes (RN) request service from RM (an RN may also share resources). Also, Dai and Wang [7] showed that the links and nodes in each grid service form a spanning tree. They define the resource-spanning tree (RST) as a tree that starts from the requestor RN (as its root) and covers all resources that are required for the requested grid service.

An example grid system is displayed in Fig. 1. The RM are shown as single and RN are shown as double circles (G_1) in the figure. As an example grid service S_1 in the grid system in Fig. 1, assume that G_1 requests the resources R_2, R_4 and R_5. For the sake of simplicity and without sacrificing from correctness, it can be assumed that the grid component G_i shares resource R_i only. So, in this example the components G_2, G_4 and G_5 share resources R_2, R_4 and R_5 respectively.

Reliability of a grid system can be estimated by using reliabilities of services provided through the system [7]. In order to evaluate the reliability for a grid service, the links and nodes that are involved in that service should be identified. Dai and Wang previously showed that the reliability of a grid service can be estimated by using the reliabilities of minimum RSTs (MRST) [7]. In Fig. 1, although there are several RST for S_1 that include all requested resources ({G_1, G_2, G_3, G_6, G_8, G_5, G_4}, {G_1, G_2, G_3, G_6, G_8, G_9, G_4, G_4}, {G_1, G_3, G_6, G_8, G_7, G_5, G_4}, etc), only one of them is an MRST; {G_1, G_2, G_3, G_5, G_4}. Other possible spanning trees in the grid are either larger than the MRST or do not include all requested resources. Moreover, since there is no other component in this example that shares the resource R_5, it can be concluded that there is only one MRST for the grid service S_1.

There are several studies in the literature that focus on the reliability of grid systems, however many of them rely on certain assumptions [5–8,19] that will be discussed in Section 3. Dai and Wang [7] present a methodology to optimally allocate the resources in a grid system in order to maximize the grid service reliability. They use a genetic algorithm to find the optimum solution efficiently among numerous possibilities. Later Levitin and Dai [19] propose dividing grid services into smaller-size tasks and subtasks, then assigning the same tasks to different RM for parallel processing. This paper focuses on the reliabilities of MRSTs in the grid system, where the reliability of an MRST is the probability for the MRST to provide the given service. The $K2$ algorithm is used to discover the MRSTs and BN to evaluate grid service reliabilities. The next section provides information about the BN and the $K2$ algorithm.

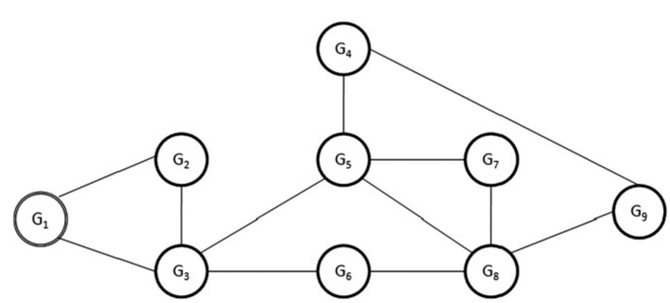

Fig. 1. A sample grid system.