

Ranking Grid-sites based on their Reliability for Successfully Executing Jobs of Given Durations

Farrukh Nadeem

Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University,
Jeddah, Saudi Arabia
Email: fabdullatif@kau.edu.sa

Abstract—Today's Grids include resources (referred as Grid-site s) from different domains including dedicated production resources, resources from university labs, and even P2P environments. Grid high level services, like schedulers, resource managers, etc. need to know the reliability of the available Grid-sites to select the most suitable from them. Modeling reliability of a Grid-site for successful execution of a job requires prediction of Grid-site availability for the given duration of job execution as well as possibility of successful execution of the job. Predicting Grid-site availability is complex due to different availability patterns, resource sharing policies implemented by resource owners, nature of domain the resource belongs to (e.g. P2P etc.), and its maintenance etc. To give a solution, we model reliability of Grid-site in terms of prediction of its availability and possibility of job success. Our availability predictions incorporate past patterns of the Grid-site availability using pattern recognition methods. To estimate possibility of job success, we consider historical traces of job execution. The experiments conducted on a trace of real Grid demonstrate the effectiveness of our approach for ranking Grid-sites based on their reliability for executing jobs successfully.

Index Terms—The Grid, Grid-site availability, Grid-site reliability, Job success rate, Reliability modeling.

I. INTRODUCTION

With the maturity of distributed technologies the composition, scale and utilization of distributed systems is continuously evolving. The large scale distributed systems like Grids and testbeds under centralized administrative controls, such as, EGEE[13], Planet Lab[5], Open Science Grid[34], Tera Grid [33], Grid'5000 [2] etc. consist of hundreds/thousands of resources [30]. These systems/environments may even include resources available for short periods, such as computers from university labs, Public Resource Computing (PRC) systems [4] and on demand resources [30]. To select the most suitable resources from such a large pool where resources vary considerably in terms of their properties, Grid high level services need to know how reliable a Grid-site is for successful execution of a job of given duration. For example, Grid resource

manager requires Grid-sites reliability to filter out and allocate suitable resources for given requests. Grid schedulers need Grid-sites reliability to map given set of activities to available resources. Performance monitoring services require Grid-sites reliability for their performance analysis.

The reliability of a Grid-site for successful execution of a job mainly depends upon availability of the Grid-site and its capacity of fault free execution of a job. The availability of a Grid-site depends upon several factors including nature of the environment the Grid-site originally belongs to, the resource owners' policies for sharing their resources in the Grid, etc. [30]. In addition, the resources may also be disconnected from the Grid environment for regular maintenance, unexpected faults, or other projects. Due to diversification of the Grid environment, complexities of the technologies involved, and the support provided by the resource owners, the jobs submitted to the Grid-sites may not execute successfully and fail. Moreover, the jobs may also fail due to several other faults [20] in the Grid environment, which may occur at any stage during job execution due to Grid middleware, software stack on the Grid-site, unavailability of required libraries, execution time restrictions from local resource manager, etc. All these factors make reliability modeling of a Grid-site for successful execution of a job a challenging problem.

To address this problem, in this paper, we model reliability of a Grid-site for successful execution of a job. To be precise we define reliability of a Grid-site as its degree to which it can accept and execute jobs successfully. Our model considers future availability of the Grid-site for the given duration of job execution as well as possibility of fault free execution of the job. The future availability of the Grid-site is predicted using a state of the art method based on Bayesian Inference. The possibility of fault free execution of a job is measured as probability of successful execution based on past submitted jobs. This is later referred as job success rate. Both of these measurements require trace data of Grid-site availability and job executions. For our present study, we considered real traces of a national Grid (the name hidden due to privacy reasons). The results of our experiments are presented in this paper. Our experiments show that considering only Grid-site availability or job success rate leads to suboptimal ranking of Grid-sites. However, considering Grid-site reliability through our

proposed model yields optimized ranking of Grid-sites.

We use the following terminology in this paper. By Grid-site we refer to a Grid resource that can accept and execute a computation request referred as a job request (or simply a job). A Grid-site is considered available if it is turned on and can accept job requests; otherwise, it is considered unavailable. Please note that we use the terms Grid-site and Grid resource interchangeably.

The rest of the paper is arranged as follows. The section II describes major characteristics of Grid-sites that are important to understand the overall reliability modeling proposed in this paper. Section III precisely describes our problem statement. Section IV describes our method of reliability modeling. The subsections IVA and IVB describe our methods for predicting Grid-site availability and assessing job success rate. Section V describes our experiments and associated findings. The related work is presented in section

VI. Finally, we conclude and describe the future work in section VII.

II. CHARACTERISTICS OF GRID-SITES AND LARGE SCALE SCIENTIFIC APPLICATIONS

Grid-sites in medium and large scale Grids are usually contributed by different organizations, yet their ownerships remain with the parent organizations. These organizations have their own policies for sharing their resources in the Grid. In general, the Grid-sites are typically characterized by their dynamic availability. Besides the resources dedicated to Grid environment, other resources are contributed for some specific durations of time (like during working hours of working days etc.). The parent organizations may remove their resources from the Grid environment for their own special needs like projects etc. Some resources are contributed in the Grid depending upon their usage by local users – the resources are available in the Grid if not in use by a local user; as soon as the local user starts working on the resource, it is taken out of the Grid. The examples of such resources include desktop computers from computer labs and P2P environments. The resources may also be taken out of the Grid for regular maintenance, updates, etc. The resources may also be unavailable due to some faults. In nutshell, Grid-sites vary largely in their availability properties that must be taken care of while selecting a Grid-site for a job execution.

Besides their computational resources and storage capacity, Grid-sites also vary with respect to available software stacks, their maintenance, software updates etc. Due to these variations and several other possible faults [20], the jobs submitted to Grid-sites may fail. Therefore, it is important to evaluate Grid-sites for their job success rate in the past.

Scientific applications are an important class of applications that require large scale computing and/or data management infrastructures like Grid [32, 29, 27, 28]. These applications usually consists of several activities/tasks with complex execution and data dependencies [22, 37, 35], referred as large scale

applications. To exploit possible parallelism of these applications, some activities are executed in parallel on multiple Grid-sites. The overall execution time of large scale scientific applications ranges from minutes to hours and even days. The efficient execution of these applications requires uninterrupted availability of several Grid-sites for a long time. However, due to dynamic availability of Grid-sites, selection of the Grid-sites that will remain available for the duration of application execution is very challenging and requires prediction of their availability.

III. PROBLEM STATEMENT

The Grid users and high level Grid middleware services, like met schedulers and resource managers, need to rank the available Grid-sites for their reliability of executing a given job successfully and then select/allocate the best Grid-site for executing the job. The reliability of a Grid-site for executing a given job successfully depends upon its availability properties as well as its rate of executing jobs successfully. The availability properties of a Grid-site can be determined from its past records of being available for different time durations. More specifically, Grid-site's availability for the duration of job execution as well as for the higher duration is of particular interest.

Ranking of the available Grid-sites requires reliability modeling of the Grid-sites. To model reliability of the Grid-sites for successful execution of given jobs, we describe the problem statement as follows.

What is the reliability (R) of a Grid-site g (Rg) that has been available for different time durations $[d_1, d_2, d_3, \dots, d_n]$ throughout its life time and have a job success rate of λg for successfully executing a job of execution time t ?

We describe our proposed approach to model Grid-site reliability in the following section.

IV. MODELING RELIABILITY OF GRID-SITES

We model reliability of a Grid-site in terms of Grid-site availability and job success rate. The future availability of the Grid-site is predicted using Bayesian Inference [31] and the job success rate is calculated from Grid-site's trace of job executions. These methods are described in detail in the following sections.

A. Grid-site Availability Predictions

We predict Grid-site availability using Bayesian Inference. Bayesian Inference estimates possibility of an event from its likelihood and prior probability (priori) conditional to its properties. Thus, Bayesian Inference by nature exploits resource availability properties by incorporating its patterns of availability from past traces as well as current behavior. In this work, we consider only two states of a Grid-site, "available" and "unavailable". We define Bayesian Inference for our two class resource state as follows. Let θ_1 and θ_2 represent the two classes to which our Grid-site state belongs. The

priori probabilities of the two states, $P(\theta_1)$ and $P(\theta_2)$, can be calculated as follows. If N is the total number of events and n_1, n_2 belong to θ_1 and θ_2 , respectively, then

$$P(\theta_1) = \frac{n_1}{N} \quad (1)$$

and

$$P(\theta_2) = \frac{n_2}{N} \quad (2)$$

Let $p(d_j|\theta_i)$ (for $i = 1, 2, j = 1, 2, \dots, n$) represents the class conditional probability density functions (PDFs), describing the distribution of feature d_j in each class. Here, the feature set D consists different time durations for which the Grid-site has been available.

$$D = [d_1, d_2, d_3, \dots, d_n]$$

These time durations are extracted in Grid-site availability characterization phase [31]. The class conditional PDFs are also calculated from the trace data during the characterization phase. Then, according to Bayesian Inference:

$$P(\theta_i | d_j) = \frac{p(d_j|\theta_i)P(\theta_i)}{p(d_j)} \quad (3)$$

where $p(d_j)$ is the PDF of d_j and in our case it can be calculated as:

$$p(d_j) = \sum_{i=1}^2 p(d_j|\theta_i)P(\theta_i) \quad (4)$$

We consider one feature from the feature set D at a time, which can have any value from its feature space. In our case, feature set only takes discrete values and thus the density functions $p(d_j | \theta_i)$ become probabilities, which are calculated as $P(d_j | \theta_i)$.

Bayesian Inference shows that we can convert the priori $P(\theta_i)$ to a posteriori (or posterior probability) $P(\theta_i | d_j)$ – the probability of the state of nature being θ_i given that feature value d_j has been used. We call $p(d_j | \theta_i)$ the likelihood of θ_i , a term chosen to indicate that other things being equal the class θ_i for which $p(d_j | \theta_i)$ is large is more likely be the true class. Bayes' formula can be expressed informally by saying that

$$posteriori = \frac{likelihood \times priori}{evidence} \quad (5)$$

Thus in our work, the availability α of a Grid-site for the time duration d_j is modeled as its posterior probability of the Grid-site for being in state "available" for the time d_j :

$$\alpha(d_j) = P(\text{available}|d_j) \quad (6)$$

B. Job Success Rate

The job success rate λ of a Grid-site g is defined as ratio of number of jobs completed successfully (N_c) on that Grid-site to the number of jobs submitted (N_s) to it.

$$\lambda_g = \frac{N_c}{N_s} \quad (7)$$

C. Reliability Model

We model reliability R of a Grid-site g available for different time durations $[d_1, d_2, d_3, \dots, d_n]$ during its whole life time and having a job successful rate λ , for successfully executing a job with execution time t as:

$$R_g(t) = \sum_{\forall d_j \geq t} \alpha(d_j) \times \lambda_g \quad (8)$$

For a Grid-site g :

$$\sum_{j=1}^n \alpha(d_j) = 1 \quad (9)$$

Thus,

$$\sum_{\forall d_j \geq t} \alpha(d_j) < 1 \quad (10)$$

Also,

$$\lambda_g \leq 1 \quad (11)$$

Therefore,

$$0 \leq R_g(t) < 1 \quad (12)$$

It means that the reliability of a Grid-site ranges between 0 and 0.99.

V. EXPERIMENTS

To demonstrate the effectiveness of the proposed approach, we modeled reliability of several Grid-sites in a national Grid (the names of the Grid and Grid-sites are anonymized due to privacy reasons). The availability of Grid-sites was monitored for a period of one year and their states (available or unavailable) were stored after every five minutes. This availability trace was used in Bayesian Inference to predict the availability. A random time (t) from the availability trace was selected and the trace data before t was used in Bayesian Inference to predict the availability of the Grid-site at time t . The probabilistic availability predicted through Bayesian Inference was then used in our availability model. The log data gathered by local job submission systems on the Grid-sites was used to calculate the job success rate at time t .

First we present the reliability (modeled through the proposed approach) of four representative Grid-sites (two from the computer labs and two Grid-sites dedicated to the Grid for production purposes). Figures 1 and 2 show the reliability of two Grid-sites from university computer labs for executing jobs of different durations (from less than one hour to 11 hours). The job success rate on these two Grid-sites was 85% and 93%, respectively. As these sites were available for small durations, their reliability decreased sharply as the duration of job execution

increased.

Figures 3 and 4 show the reliability of two dedicated Grid-sites for executing jobs of different durations (from less than one day to 11 days). These Grid-sites were meant to be available all the times for production purposes. Although the dedicated Grid-sites remained available for more than 11 days, for simplicity reasons we have shown data for 11 days only. The job success rates on these two sites were 98% and 99%, respectively. Since these Grid-sites were mostly available for longer time durations, their reliability decreased at a smaller rate than that of Grid-sites from the Computer labs.

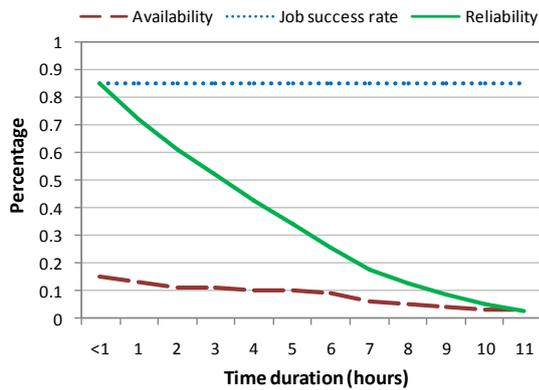


Fig. 1. Reliability of a Grid-site 1 (from computer labs) for executing jobs of different durations

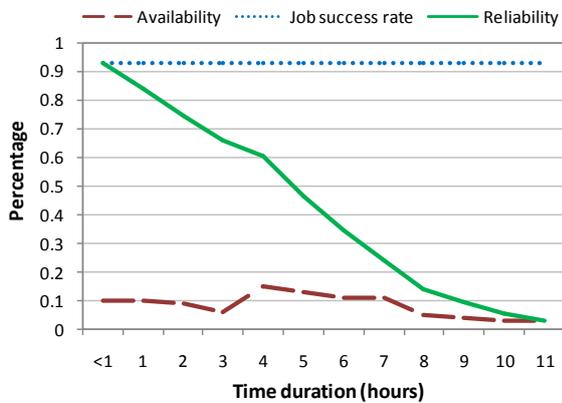


Fig. 2. Reliability of a Grid-site 2 (from computer labs) for executing jobs of different durations.

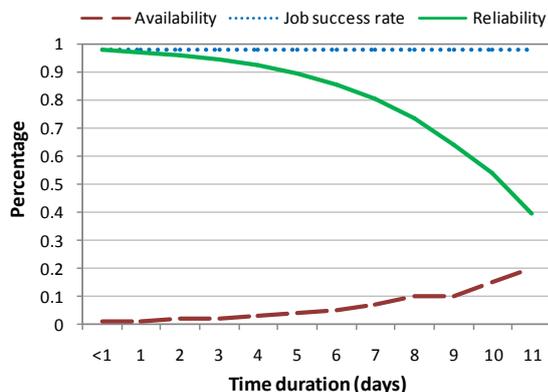


Fig. 3. Reliability of a Grid-site 3 (production site dedicated for Grid users) for executing jobs of different durations.

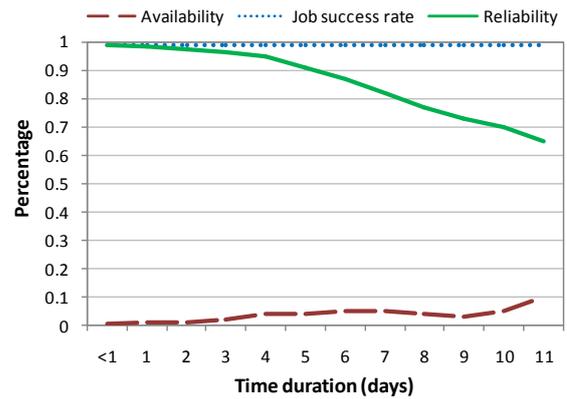


Fig. 4. Reliability of a Grid-site 4 (production site dedicated for Grid users) for executing jobs of different durations.

Next, we compared 20 Grid-sites at particular time for their job success rate, availability and reliability for a job of 30 minutes duration. Their respective rankings are shown in figures 5, 6 and 7. It can be observed that the Grid-sites ranked differently when the criteria (job success rate or availability or reliability) for ranking the Grid-sites varied. It can also be observed that considering one factor at a time (i.e. job success rate or availability) will result in a different ranking and thus will be suboptimal. However, ranking based on reliability (that considers both availability and job success rate) results in optimal order.

VI. RELATED WORK

There have been several efforts for quantifying the reliability of resources in distributed systems, like in [36, 25, 17, 3]. These efforts are mainly based on architectural considerations of distributed systems. However, it is very difficult to compute reliability of arbitrary structured distributed systems [24]. Xie et al. [38] modeled reliability of hardware systems, software, integrated systems, distributed computing systems and Grid computing systems.

Xie et al. [38] also modeled overall reliability of Grid computing systems based on known failure rates of nodes, links and other known components. Later Dai et al. [6] refined the work by Xie et al. [38] by considering a hierarchical model of the Grid. The authors used Markov chain to model probabilities of different failures. In their work Dai et al. [6] modeled reliability of collection of Grid resources communicating on shared links. Later in [8, 7, 9], the authors modeled reliability and performance of services in a tree structured Grid using graph theory and probability theory. The authors also extended their work for optimal allocation of Grid resources to maximize the reliability of services in a tree structured Grid. Graph theory based methods have also been employed for reliability assessment by Dakil et al. [10]. Levitin et al. [26] modeled reliability and performance of services in Grid computing systems with star architectures. However, these work mainly considered job failure rate of Grid resources/services.

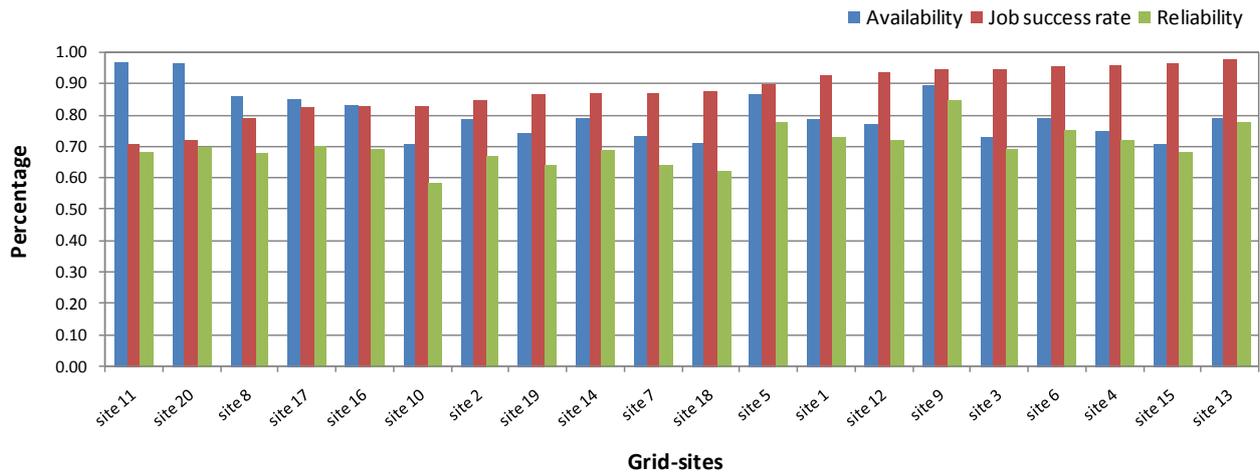


Fig. 5. Ranking of 20 Grid-sites for their job success rate in the Grid. (The right-most Grid-site has the highest rank).

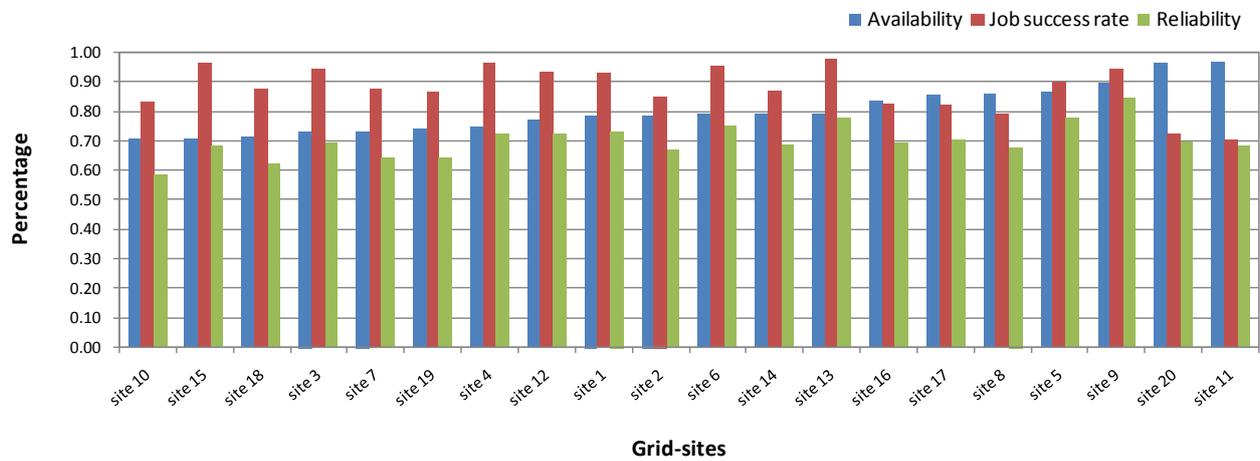


Fig. 6. Ranking of 20 Grid-sites for their availability in the Grid. (The right-most Grid-site has the highest rank).

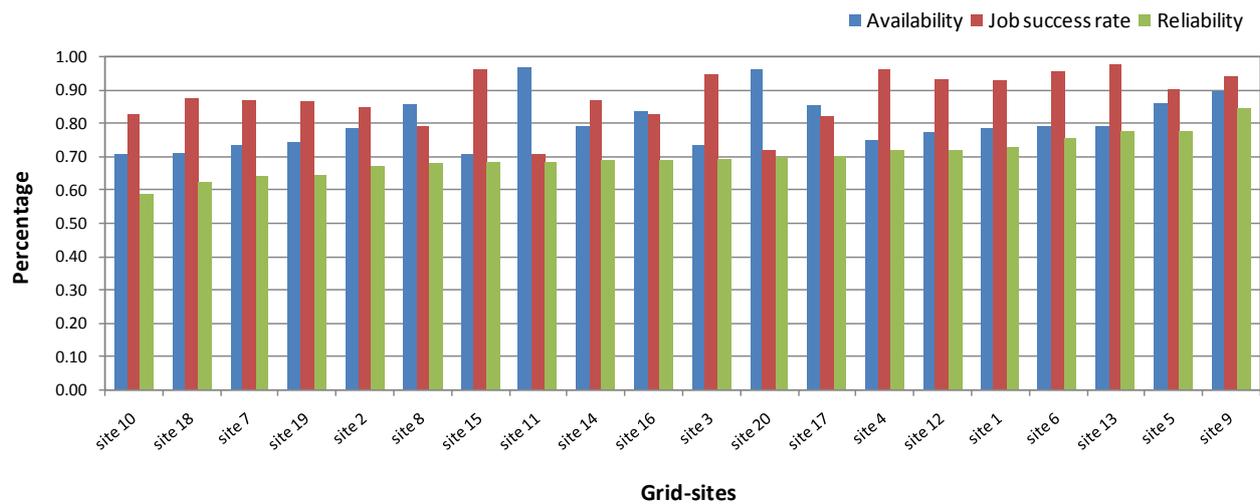


Fig. 7. Ranking of 20 Grid-sites for their reliability in the Grid. (The right-most Grid-site has the highest rank).

Gholami et al. [15] proposed a Monte Carlo simulation based algorithm to estimate reliability of Grid programs and the whole Grid system. To estimate the reliability of involved nodes and links, their algorithm considers rate

of data transmission through the links and the processing time on nodes. Later, these authors improved their work to maximize the reliability of Grid computing systems using Simulated Annealing [16]. The authors in [14]

proposed a Graph theory and Bayesian networks based algorithm to estimate reliability and latency of Grid computing systems. Like others [1], the authors also consider rate of data transfer through the links among Grid nodes. Beside others, Kuar [23] and Gran et al. [18] also used Bayesian networks to estimate reliability. Doguc et al. [12] proposed a holistic method for estimating system reliability using Bayesian networks. The authors employed K2 data mining algorithm to find associations between the system components. In their later work [11], the authors further improved their work by making their approach independent of any prior knowledge about the grid system structure. The authors used K2 algorithm to automatically discover the structure of Grid system from historical data and to find minimum resource spanning trees (MRST) in the Grid. In next step, the authors used Bayesian networks to model the MRST and estimate reliability of the Grid. Guo et al. [19] modeled reliability of Grid services based on Local Node Fault Recovery (LNFR) mechanism and demonstrated the influence of fault recovery on Grid service reliability. Hussin et al. [21] modeled reliability in resource management for distributed systems through adaptive reinforcement learning. The methodology of reinforcement learning is used in conjunction with neural network to help Grid scheduler adapt to dynamic changes in Grid environment.

In contrast, our focus is on modeling reliability of Grid-sites for successful execution of Grid jobs. To model Grid-site reliability, we consider Grid-sites' availability for the duration of job execution as well as previous job success rate on the Grid-site. To the best of our knowledge, we are the first to rank the Grid-sites based on their future availability for the duration of job execution as well as their job success rate. This reliability measure is helpful for Grid users, meta schedulers and resource managers in optimized ranking of Grid-sites for their possibility of executing a job successfully. We have also demonstrated through our experiments that considering either job success rate or availability results in sub-optimal selection of Grid-sites.

VII. CONCLUSION AND FUTURE WORK

Grid schedulers, resource manager and performance analysis services require Grid-site s' ranking based on their reliability for successful execution of Grid jobs. Modeling reliability of a Grid-site for successful execution of a job requires predictions about its availability for the duration of job execution as well as its job success rate. Effective predictions of Grid-site availability requires past patterns of resource availability. We predict Grid-site availability using methods from pattern classification and recognition, which incorporate resource past behavior. The job success rate is calculated as probability of successful job executions using trace data. Our method can effectively rank Grid-sites to optimize reliability of Grid-sites for successful execution of jobs. In future, we plan to evaluate our reliability model using other methods of resource availability

predictions. We also plan to develop methods to cross verify our reliability measurements.

REFERENCES

- [1] Waseem Ahmed and Yong Wei Wu. A survey on reliability in distributed systems. *Journal of Computer and System Sciences*, 79(8):1243–1255, 2013.
- [2] ALADDING5K, INRI. Grid'5000 project, 2014. <https://www.grid5000.fr/>.
- [3] Raid Alsoghayer and Karim Djemame. Resource failures risk assessment modelling in distributed environments. *Journal of Systems and Software*, 88(0):42 – 53, 2014.
- [4] David P. Anderson. Boinc: A system for public resource computing and storage. In *IEEE/ACM International Workshop on Grid Computing*, Washington, DC, USA, 2004.
- [5] Brent Chun, David Culler, Timothy Roscoe, Andy Bavier, Larry Peterson, Mike Wawrzoniak, and Mic Bowman. Planet lab: An overlay tested for broad coverage services. *SIGCOMM Computation and Communication Review*, 33(3):3–12, July 2003.
- [6] Yuan Shun Dai and G. Levitin. Reliability and performance of tree structured grid services. *IEEE Transactions on Reliability*, 55(2):337–349, June 2006.
- [7] Yuan Shun Dai and Gregory Levitin. Optimal resource allocation for maximizing performance and reliability in tree structured grid services. *IEEE Transactions on Reliability*, 56(3):444–453, Sept 2007.
- [8] Yuan Shun Dai, Gregory Levitin, and Kishor S. Trivedi. Performance and reliability of tree structured grid services considering data dependence and failure correlation. *IEEE Transactions on Computers*, 56(7):925–936, July 2007.
- [9] Yuan Shun Dai, Yi Pan, and Xukai Zou. A hierarchical modeling and analysis for grid service reliability. *IEEE Transactions on Computers*, 56(5):681–691, May 2007.
- [10] Manal Dakil, Christophe Simon, and Taha Boukhobza. Connectivity condition for structural properties using a graph theoretical approach: probabilistic reliability assessment. *System, Structure and Control*, 5(1):72–77, 2013.
- [11] Ozge Doguc and Jose Emmanuel Ramirez Marquez. An automated method for estimating reliability of grid systems using Bayesian networks. *Reliability Engineering & System Safety*, 104(0):96 – 105, 2012.
- [12] Ozge Doguc and Jose Emmanuel Ramirez Marquez. A generic method for estimating system reliability using Bayesian networks. *Reliability Engineering & System Safety*, 94(2):542 – 550, 2009.
- [13] EGEE consortium. Enabling Grids for science (EGEE), 2010. <http://euegeeorg.web.cern.ch/euegeeorg/index.html>.
- [14] Mohana Farzin, Poorya Khodabande, and Hadi Toofani. Reliability and latency calculation in grid computing systems. In *5th International Conference on Application of Information and Communication Technologies (AICT)*, pages 1–6, Oct 2011.
- [15] Ehsan. Gholami, Amir Masood Rahmani, and Ahmad Habibzad Navin. Using Monte Carlo simulation in grid computing systems for reliability estimation. In *Eighth International Conference on Networks, 2009 (ICN '09)*, pages 380–384, March 2009.
- [16] Ehsan Gholami, Amir Masoud Rahmani, and Reza Farshidi. Using simulated annealing to improve reliability of grid computing systems. In *The Fourth International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP 2010)*, pages 17–22, 2010.

- [17] Katerina Goševa Popstojanova and Kishor S Trivedi. Architecture based approaches to software reliability prediction. *Computers & Mathematics with Applications*, 46(7):1023 – 1036, 2003.
- [18] Bjørn Axel Gran and Atte Hemline. A Bayesian belief network for reliability assessment. In *Computer Safety, Reliability and Security*, pages 35–45. Springer, 2001.
- [19] Suchang Guo, Hong Zhong Huang, Zhonglai Wang, and Min Xie. Grid service reliability modeling and optimal task scheduling considering fault recovery. *IEEE Transactions on Reliability*, 60(1):263–274, 2011.
- [20] Juergen Hofer and Thomas Fahringer. A multiperspective taxonomy for systematic classification of grid faults. In *Proceedings of the 16th Euromicro Conference on Parallel, Distributed and Network Based Processing, PDP'08*, pages 126–130, Washington, DC, USA, 2008. IEEE Computer Society.
- [21] Masnida Hussin, Nor Asilah Wati Abdul Hamid, and Khairul Azhar Kasmiran. Improving reliability in resource management through adaptive reinforcement learning for distributed systems. *Journal of Parallel and Distributed Computing*, 31(10):60–68, 2014.
- [22] Institute of Physical and Theoretical Chemistry, TU Vienna. WIEN2k: An Augmented Plane Wave plus Local Orbitals Program for Calculating Crystal Properties. <http://www.wien2k.at/>, 2014.
- [23] Ipneet Kaur. Estimating Grid Reliability Using Bayesian Networks. PhD thesis, THAPAR UNIVERSITY, 2011.
- [24] V K P Kumar, S Hariri, and C S Raghavendra. Distributed program reliability analysis. *IEEE Transactions on Software Engineering*, 12(1):42–50, January 1986.
- [25] Way Kuo and V.R. Prasad. An annotated overview of system reliability optimization. *Reliability*, *IEEE Transactions on*, 49(2):176–187, Jun 2000.
- [26] Gregory Levitin, Yuan Shun Dai, and Hanoeh Blenheim. Reliability and performance of star topology grid service with precedence constraints on subtask execution. *IEEE Transactions on Reliability*, 55(3):507–515, Sept 2006.
- [27] Farrukh Nadeem and Thomas Fahringer. Predicting the execution time of grid workflow applications through local learning. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, SC '09*, pages 1–12, New York, NY, USA, Nov 2009. ACM.
- [28] Farrukh Nadeem and Thomas Fahringer. Optimizing execution time predictions of scientific workflow applications in the grid through evolutionary programming. *Future Gener. Comput. Syst.*, 29(4):926–935, June 2013.
- [29] Farrukh Nadeem, Radu Prodan, and Thomas Fahringer. Optimizing performance of automatic training phase for application performance prediction in the grid. In *Proceedings of Third International Conference on High Performance Computing and Communications, HPCC'2007*, pages 309– 321, Houston, USA, September 26-28 2007. Springer.
- [30] Farrukh Nadeem, Radu Prodan, and Thomas Fahringer. Characterizing, modeling and predicting dynamic resource availability in a large scale multipurpose grid. In *8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2008)*, pages 348–357. IEEE Computer Society, 2008.
- [31] Farrukh Nadeem, Radu Prodan, Thomas Fahringer, and Vincent Keller. An evaluation of availability comparison and prediction for optimized resource selection in the grid. In *From Grids to Service and Pervasive Computing*, pages 63–76. Springer US, 2008.
- [32] Farrukh Nadeem, Muhammad Murtaza Yousaf, Radu Prodan, and Thomas Fahringer. Soft benchmarks based application performance prediction using a minimum training set. In *E-SCIENCE '06: Proceedings of the Second IEEE International Conference on science and Grid Computing*, page 71, Amsterdam, Netherlands, December 2006.
- [33] National Science Foundation. The TeraGrid Project, 2010. <http://www.teragrid.org/>.
- [34] OSG Consortium. Open Science Grid, 2014. <http://www.opensciencegrid.org/>.
- [35] Felix Schueller, Jun Qin, Farrukh Nadeem, Radu Prodan, Thomas Fahringer, and Georg Mayr. Performance, Scalability and Quality of the Meteorological Grid Workflow MeteoAG. In *2nd Austrian Grid Symposium, Innsbruck, Austria. OCG Verlag*, September 21-23 2006.
- [36] Daniel P. Siewiorek and Robert S. Swarz. *Reliable Computer Systems: Design and Evaluation* (3rd Ed.). A K Peters, Ltd., Natick, MA, USA, 1998.
- [37] Dieter Theiner and Marek Wiecek. Reduction of calibration time of distributed hydrological models by use of grid computing and nonlinear optimization algorithms. In *Proceedings of the 7th International Conference on Hydro informatics (HIC 2006)*, September 2006.
- [38] Min Xie, Yuan Shun Dai, and Kim Leng Poh. *Computing System Reliability: Models and Analysis*. Springer US, 2004.

Authors' Profiles



Farrukh Nadeem received his Ph.D. in 2009 from the University of Innsbruck, Austria. Since July 2009, he has been working as an Assistant Professor. His main research interests include performance modeling and prediction, and scheduling scientific workflows in distributed systems, particularly the Grid and the Cloud. He has been involved in a several Austrian and Saudi research and development projects. Farrukh is author of more than 22 papers, including four book chapters. He holds several distinctions and awards during his educational career.

How to cite this paper: Farrukh Nadeem, "Ranking Grid-sites based on their Reliability for Successfully Executing Jobs of Given Durations", *IJCNIS*, vol.7, no.5, pp.9-15, 2015. DOI: 10.5815/ijcnis.2015.05.02