Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Future Generation Computer Systems 23 (2007) 737–747
FGCS
FUTURE GENERATION COMPUTER SYSTEMS
www.elsevier.com/locate/fgcs

Analytical communication networks model for enterprise Grid computing

Bahman Javadi[a], Mohammad K. Akbari[a], Jemal H. Abawajy[b,*]

[a] Amirkabir University of Technology, Computer Engineering and Information Technology Department, 424 Hafez Ave., Tehran, Iran
[b] Deakin University, School of Engineering and Information Technology, Geelong, VIC. 3217, Australia

Received 22 April 2006; received in revised form 23 November 2006; accepted 24 November 2006
Available online 2 April 2007

Abstract

This paper addresses the problem of performance analysis based on the communication modeling of large-scale heterogeneous distributed systems, with an emphasis on enterprise Grid computing systems. The study of communication layers is important, as the overall performance of a distributed system often critically hinges on the effectiveness of this part. We propose an analytical model that is based on probabilistic analysis and queuing networks. The proposed model considers the processor as well as network heterogeneity of the enterprise Grid system. The model is validated through comprehensive simulations, which demonstrate that the proposed model exhibits a good degree of accuracy for various system sizes, and under different working conditions.

Keywords: Enterprise Grid; Performance analysis; Analytical modeling; Heterogeneity

1. Introduction

Advances in computational and communication technologies have made it economically feasible to conglomerate multiple independent clusters, leading to the development of the large-scale distributed systems commonly referred to as Grid computing [1–3]. Grids can be classified in two ways, according to their architecture and presence: global Grids and enterprise Grids [30]. These two categories have varying characteristics and are suitable for different scenarios. Global Grids are established over the public Internet, are characterized by a global presence, comprise highly heterogeneous resources, present more sophisticated security mechanisms, focus on single sign-ons, and are mostly batch-job oriented [30]. In contrast, enterprise Grid computing systems consist of resources spread across an enterprise, and provide services to users within that enterprise and are managed by a single organization [4,5,29]. They can be deployed within large corporations that have a global presence even though they are limited to a single enterprise [13]. Organizations may also want to go beyond their Grids to share resources with new partners when their applications require computing resources that surpass what their own Grids can offer [29]. In this way, by using the extra resources offered by other partners, they can improve their performance as well as increasing their agility.

This paper addresses the problem of performance analysis based on the communication modeling of large-scale heterogeneous distributed systems, with an emphasis on enterprise Grid computing systems. We are motivated to study this problem for a number of reasons. First, interconnection network design plays a central role in the design and development of enterprise Grid computing. Second, due to interconnection network's contention problems [6], having a fast communication network does not necessarily guarantee a good performance from the enterprise Grid computing system built on it. The contention problems, which adversely affect the overall performance, can happen in *host nodes, network links*, and *network switches* [6]. Node contention happens when multiple data packets compete to contain a receive channel of a node, and link contention occurs when two or more packets share a communication link. Switch contention is due to unbalanced traffic flow through the switch, which can result in an overflow of the switch buffer.

The contribution of this paper is to propose an analytical model for the communication networks of enterprise Grid systems. The proposed model is based on probabilistic analysis and queuing networks to analytically evaluate the

* Corresponding author. Tel.: +61 3 5227 1376; fax: +61 3 5227 2028.
E-mail addresses: javadi@ce.aut.ac.ir (B. Javadi), akbari@ce.aut.ac.ir (M.K. Akbari), jemal@deakin.edu.au (J.H. Abawajy).

0167-739X/$ - see front matter
doi:10.1016/j.future.2006.11.002

performance of communication networks for enterprise Grid systems. The model takes into account processor as well as network heterogeneity among clusters. The model is validated through comprehensive simulations, which demonstrate that the proposed model exhibits a good degree of accuracy for various system sizes and under different operating conditions. Several analytical performance models of multi-computer systems have been proposed in the literature for different interconnection networks and routing algorithms (e.g., [7–12]). However, analytical models for the systems of greatest interest are generally rare, and most of the existing works are based on homogenous single cluster systems [13–16] with the exception of [17], which looked at processor heterogeneity. Moreover, in [18] a queuing model based on input and server distributions was proposed to analyze a special Grid system, VEGA 1.1. The majority of these works are based on job level modeling, but in contrast, we intend to propose an analytical model of the communication layer of enterprise Grid systems to provide a more accurate performance analysis and prediction. To the best of our knowledge, our work is the first which deals with heterogeneous enterprise Grid environments.

The rest of the paper is organized as follows. In Section 2, a brief description of the enterprise Grid system used in this paper is presented. In Section 3, we give a detailed description of the proposed analytical communication model. In Section 4, we present an experimental evaluation to validate the proposed model. Finally, we summarize our findings and conclude the paper in Section 5.

2. System description

The computational Grid architecture which is used in this paper is shown in Fig. 1. The system is composed of C clusters, each cluster is composed of $N_i, i \in \{0, 1, \ldots, C-1\}$, computing nodes, each node with τ_i processing power, and its associated memory module. In our model, we express the processing power of the various processors in each cluster relative to a fixed reference processor [17], and not relative to the fastest processor, which is used in the most works on heterogeneous parallel systems. Although the latter choice may appear more natural, since it makes it possible to obtain the speedup by comparing the performance of the parallel system with that of the fastest single node available, we think that choosing a fixed reference allows a clearer performance analysis, especially if we vary the number and/or the power of nodes. So the relative processing power of each node can be stated as $s^{(i)} = \tau_i/\tau_f$, where f is the reference machine number. Since we consider processor heterogeneity among clusters, the total relative processing power and the average relative processing power in the system are as follows, respectively:

$$S = \sum_{i=0}^{C-1} s^{(i)} \tag{1}$$

$$\bar{s} = \frac{S}{C}. \tag{2}$$

Each cluster in the enterprise Grid system has two communication networks: an Intra-Communication Network

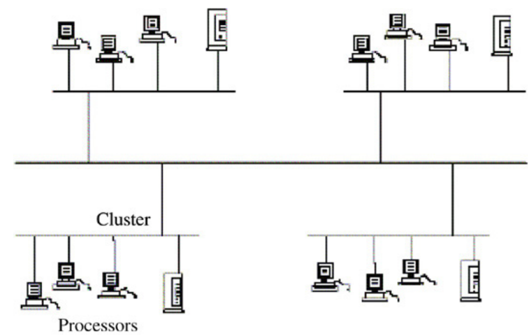

Fig. 1. Enterprise Grid computing architecture.

(ICN1) and an intEr-Communication Network (ECN1). The ICN1 is used for the purpose of passing messages between processors within a cluster, while the ECN1 is used to transmit messages between clusters and for the management of the entire system. To interconnect clusters, the ECN1 is connected through a set of concentrators/dispatchers [19] to the external network, i.e., ICN2. Since high performance computing clusters typically utilize *Constant Bisectional Bandwidth* networks [20–22], we adopted m-port n-tree [23] as fixed arity switches to construct the topology for each cluster. An m-port n-tree topology consists of N_{pn} processing nodes and N_{sw} communication switches, which can be calculated with following equations:

$$N_{pn} = 2 \times \left(\frac{m}{2}\right)^n \tag{3}$$

$$N_{sw} = (2n-1) \times \left(\frac{m}{2}\right)^{n-1}. \tag{4}$$

In addition, each communication switch itself has m communication ports $\{0, 1, 2, \ldots, m-1\}$ that are attached to other switches or processing nodes. Every switch (except the root switches) uses ports in the range of $\{0, 1, 2, \ldots, (m/2)-1\}$ to make connections with its descendants or processing node, and uses ports in the range of $\{(m/2), (m/2)+1, \ldots, m-1\}$ for connection with its ancestors. This means that the internal switches have $m/2$ up links and $m/2$ down links connected to their ancestors and descendants respectively. It can be shown that the m-port n-tree is a full bisection bandwidth topology [24], so link contention doesn't occur in such network.

In this paper, we used wormhole flow control and deterministic routing, which are commonly used in cluster network technologies such as Myrinet, Infiniband and QsNet [25]. We used a deterministic routing based on Up*/Down* routing [26], which is proposed in [24]. In this algorithm, each message experiences two phases, an *ascending phase* to get to a Nearest Common Ancestor (NCA), followed by a *descending phase*. Furthermore, since this algorithm performs a balanced traffic distribution, the switch contention problem will not be present [24].

3. The analytical communication model

In this section, we develop the analytical communication model for the enterprise Grid system discussed in the previous