Chapter 9

Adaptive Control of Redundant Task Execution for Dependable Volunteer Computing

Hong Wang
Tohoku University, Japan

Yoshitomo Murata
Tohoku University, Japan

Hiroyuki Takizawa
Tohoku University, Japan

Hiroaki Kobayashi
Tohoku University, Japan

ABSTRACT

On the volunteer computing platforms, inter-task dependency leads to serious performance degradation for failed task re-execution because of volatile peers. This paper discusses a performance-oriented task dispatch policy based on the failure probability estimation. The tasks with the highest failure probabilities are selected for dispatch when multiple task enquiries come to the dispatcher. The estimated failure probability is used to find the optimized task assignment that minimizes the overall failure probability of these tasks. This performance-oriented task dispatch policy is evaluated with two real world trace data sets on a simulator. Evaluation results demonstrate the effectiveness of this policy.

INSTRUCTION

Volunteer computing (Anderson, 2004) uses Internet-connected individual computers to solve computing problems. The pioneering research projects, including GIMPS (The Great Internet Mersenne Prime Search, http://www.mersenne.org), SETI@home (Anderson, 2004) and Distributed.net (http://www.distributed.net) are rather successful. GIMPS has already found a total of 9 Mersenne primes, each of which was the largest known prime number at the time of discovery. SETI@home has identified several candidate spots for extraterrestrial intelligence. Distributed.
net has successfully provides the solutions of the DES, RC5-32/12/7 ("RC5-56"), and RC5-32/12/8 ("RC5-64") of the RSA secret-key challenge.

Nowadays, there are several well-known volunteer computing platforms such as Folding@home (http://folding.stanford.edu), BOINC (Berkeley Open Infrastructure for Network Computing, http://boinc.berkeley.edu), XtremWeb (Cappello, 2005), Entropia (Chien, 2003), Alchemi (Luther, 2005), and JNGI (Verbeke, 2005) to name a few. The volunteer computing platforms are providing more computing power than any supercomputers, clusters, or grid, and the disparity will grow over time. It is because of a large number of Internet-connected personal computers and latest generation game consoles. By November 2010, the most powerful volunteer computing platform - Folding@home achieved about 4 Petaflops computing power by connecting more than 5,700,000 CPUs (http://fah-web.stanford.edu/cgi-bin/main.py?qtype=osstats). In contrast, the fastest supercomputer, Tianhe-1A achieves 2.566 Petaflops for the high-performance LINPACK benchmark (http://www.top500.org).

Despite the massive computing power offered by the existing volunteer computing platforms, they are lacking support for inter-task dependency. Our previous work solved this issue with a workflow management mechanism (Wang, 2007). However, inter-task dependency results in a status that none of the un-dispatched tasks can be dispatched, because these un-dispatched tasks require the results of one or several of the tasks that are being executed. This status may lead to serious performance degradation, because of the frequent task failures of volatile peers in volunteer computing platforms. Therefore, a redundant task dispatch policy (Wang, 2007) has been proposed to mitigate the performance degradation. Although the redundant task dispatch policy shown a significant performance improvement compared to the non-redundant one, it has a major limitation: the average failure rate model is not the best fit for the volunteer peers in the real world. Thus, this paper extends the policy so as to address the limitation.

This paper discusses a performance-oriented task dispatch policy for volunteer computing platforms. A heuristics-based mechanism for failure probability estimation is proposed based on a life cycle model of volunteer peers and the statistical data. The tasks with the highest failure probabilities are dispatched when multiple task enquiries come to the dispatcher. The estimated failure probability is used to find the optimized task assignment that minimizes the overall failure probability of these tasks. Once the optimized assignment is found, the dispatched tasks are sent to the workers. At the same time, the failure probabilities and other runtime information of the tasks are updated. While multiple types of workers exist in the real world, their different availability characteristics have to be considered. Thus, this work also studies the performance impact of identifying multiple worker types.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 proposes a heuristics-based failure probability estimation method. Section 4 introduces the design of the least failure probability dispatch policy. Section 5 evaluates the proposed policy using a simulator, in terms of the total process time. Section 6 concludes and summarizes this paper.

**RELATED WORK**

The failure probability is estimated based on the analysis of peer availability data. The resource availability problem has been studied a lot for clusters, servers, PCs in a corporate network, grid, and volunteer computing systems.
Related Content

MaGate: An Interoperable, Decentralized and Modular High-Level Grid Scheduler
[www.igi-global.com/article/magate-interoperable-decentralized-modular-high/46048?camid=4v1a](www.igi-global.com/article/magate-interoperable-decentralized-modular-high/46048?camid=4v1a)

Toward an Access Infrastructure for Mobile Cloud: A Channel Assignment Scheme for Wireless Mesh Networks
[www.igi-global.com/article/toward-an-access-infrastructure-for-mobile-cloud/95115?camid=4v1a](www.igi-global.com/article/toward-an-access-infrastructure-for-mobile-cloud/95115?camid=4v1a)

Molecular Structure Determination on the Grid
[www.igi-global.com/chapter/molecular-structure-determination-grid/64519?camid=4v1a](www.igi-global.com/chapter/molecular-structure-determination-grid/64519?camid=4v1a)

Scenarios of Next Generation Grid Applications in Collaborative EnvironmentsA Business-Technical Analysis
[www.igi-global.com/chapter/scenarios-next-generation-grid-applications/19338?camid=4v1a](www.igi-global.com/chapter/scenarios-next-generation-grid-applications/19338?camid=4v1a)