Sabotage-Tolerant High-Performance Volunteer Computing Based on Credibility

著者	渡邊 寛
号	16
学位授与機関	Tohoku University
学位授与番号	情博第503号
URL	http://hdl.handle.net/10097/59902

氏名（本籍地）	渡邊　寛（わたなべ　かん）
学位の種類	博　士（情報科学）
学位記番号	情　博　第　５０３　号
学位授与年月日	平成２３年　３月２５日
学位授与の要件	学位規則第４条第１項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）情報基礎科学専攻
学位論文題目	Sabotage-Tolerant High-Performance Volunteer Computing Based on Credibility（信頼度評価に基づく耐妨害性を有する高性能ボランティアコンピューティングに関する研究）
論文審査委員	（主査）東北大学教授　亀山　充隆 東北大学教授　木下　哲男　　東北大学教授　篠原　歩 東北大学教授　小林　広明

論文内容の要旨

Chapter 1. Introduction

Volunteer computing (VC) is an Internet-based parallel computing paradigm that enables any Internet participant to contribute the idle computing resources of their desktop PCs (such as CPU cycles and storage) to efforts aimed at solving large parallel problems. While there has been rapidly growing interest in VC as a ultrahigh performance computing environment, VC still has a mandatory issue for reliable computing, that is, the resources in VC may behave erratically due to hardware/software failures or virus infection, or may behave maliciously to falsify the computation, each of which results in sabotage to the computation.

Against the sabotaging, some sabotage-tolerance mechanisms are proposed for reliable computation. However, the sabotage tolerant performance and the drawbacks of those mechanisms have been hidden in real VC systems since they are proposed and evaluated based on unrealistic VC models. The objective of this research is to establish a sabotage-tolerant and high-performance volunteer computing platform, focusing on solving the following major issues with the existing sabotage-tolerance mechanisms.

- Inefficient redundant computation (excess redundancy)
- Incapable of various environments (presence of unknown parameters)
- Incapable of particular sabotaging (imperfections of checking techniques)

Chapter 2. Volunteer Computing Platform

This chapter provides some background knowledge of VCs. As described in this chapter, many VC projects have run successfully or are currently running on the Internet. Those projects are developed using a VC framework, which provides templates and simplifies the work of creating a VC project. Nowadays it is not so difficult for any scientist to use VC for high performance computing.

While VC becomes easy to use, it still has a mandatory issue for reliable computing; that is, saboteurs in VC systems may behave maliciously and return incorrect results, each of which degrades computational correctness. Against these sabotaging, current VC frameworks provide some sabotage-tolerance mechanisms such as credibility-based voting to decrease error rate of a computation and improve the computational reliability of VC systems. However, those mechanisms are proposed and evaluated in unrealistic VC models such as the simple random attack model without workers' defection. The performance and drawbacks of those mechanisms are hidden in real VC systems. Therefore, current VC frameworks such as BOINC provides the simplest sabotage-tolerance mechanism based on a redundant computation, which significantly degrades the performance instead of ensuring the reliability of computation. This motivates us to reveal the sabotage-tolerance performance and drawbacks of current sabotage-tolerance mechanisms, and improve them for higher performance and more reliable VC systems.

Chapter 3. Credibility-Based Mechanism and Job Scheduling

Since the presence of saboteurs and its adverse effects becomes known, many researchers aim to provide "sabotage-tolerance mechanisms", which decreases error rate of a computation and improves the computational reliability of VC systems. However, those mechanisms are proposed and evaluated in unrealistic VC models such as the simple random attack model.

In this chapter, these sabotage-tolerance mechanisms such as credibility-based mechanisms are introduced and evaluated in more realistic models to study the sabotage-tolerance performance and drawbacks. This study indicates that the choices of sabotage-tolerance mechanisms and job scheduling methods have a significant impact on the performance.

Then, for more efficient sabotage tolerance and higher performance of VC systems, a dynamic job scheduling method is proposed. The key factor to reduce the computation time is how to save unnecessary job allocations. It is, however, difficult to find during a computation which results will be unnecessary at the end of the computation because the credibility of jobs changes dynamically depending on the unpredictable results of jobs returned for normal and spotter jobs. To predict this, two novel metrics are defined in the proposed method: the expected-credibility and the expected number of results. Using these two metrics, the proposed scheduling method can select a proper job to be executed prior to others, thereby achieving a reduction in the computation time of a computation.

Chapter 4. Optimal Spot-checking for High-Performance Volunteer Computing

As shown in Chapter 3, credibility-based voting is a promising approach to the sabotage tolerance of VC systems. However, the same problem as for the simple voting methods remains unsolved. That is, it relies on a project owner to determine the spot-check rate. The degree of redundant jobs is an important parameter which impacts on the performance of the system. A large rate could result in an excessive checking, which increases the computation time due to the spot-checking itself. Therefore, the estimation of the optimal spot-check rate is the most important issue to realize high-performance and high-reliable VC systems.

In this chapter, a technique is proposed for estimating the optimal spot-check rate to minimize the computation time of the credibility-based voting. The key idea for the estimation is to represent the mathematical expectation of the computation time as a function of spot-check rate q. The main contributions of this work are the follows. (1) Developing an optimization technique for spot-checking on the credibility-based voting with ENR-based job selection. This technique allows project owners to choose an appropriate spot-check rate to minimize the computation time of VC projects. (2) Verifying the accuracy of estimated spot-check rate through a custom VC simulator and showing that the estimated spot-check rate always matches well the optimal rate. (3) Revealing the performance of several job selection methods using the estimated spot-check rate and pointing out the critical problem of an arbitrary chosen spot-check rate.

Chapter 5. Generalized Spot-checking for Sabotage-Tolerant Volunteer Computing

In spot-checking-based sabotage-tolerance methods such as credibility-based voting, the results of checking are utilized in the estimation of sabotaging frequency and the calculation of credibility. Those methods work well as long as the saboteurs never distinguish spotter jobs; that is, they implicitly assume that the results of spot-checking are fully-reliable. However, generating such indistinguishable spotter jobs is still an open and tough problem because it requires a huge number of reliable nodes or computation time to prepare a number of various spotter jobs and the correct results, which are impractical in real VC systems. Thus, the result of spot-checking cannot be fully-reliable in real VCs. Saboteurs may return correct results only for spotter jobs, while sabotaging normal jobs to disturb the computation. This makes the spot-checking-based methods useless for guaranteeing the computational correctness of VCs.

This chapter introduces an idea of generalized spot-checking to guarantee the computational correctness under the situation that the result of spot-checking is not fully-reliable. The main contributions of this work are the follows. (1) Developing a generalized formula of the credibility by introducing the probability c that represents the accuracy of spot-checking. This generalization allows credibility-based voting to guarantee the computational correctness even if saboteurs distinguish spotter jobs. Through simulations, the accuracy of the generalized formula is verified. (2) Enabling to apply a check-by-voting method to credibility-based voting to improve the performance of VCs, while guaranteeing the computational correctness. This application has become possible because of the generalized formula. Through simulations, it is compared the performance of VCs with and without check-by-voting.

Chapter 6. Group-based Job Scheduling for Time Limited Volunteer Computing

Credibility-based voting with check-by-voting proposed in Chapter 5 becomes a promising approach to high-performance and reliable VC in that the computational correctness is guaranteed with little redundant computation. However, the basic job scheduling methods for credibility-based voting do not work well for check-by-voting. Especially, for VCs with severe deadlines, those methods significantly degrade the performance of VC systems due to the presence of "half-finished jobs". The number of half-finished jobs has a significant impact on performance because they do not increase the throughput of VC systems, while wasting the computing resources. Moreover, such half-finished jobs prevent check-by-voting from checking the credibility of participants sufficiently. Since the deadline is given based on numerous factors such as the availability of VC systems and the demand of computation projects, a novel job scheduling method is necessary for credibility-based voting with check-by-voting to support various VCs, especially those with severe deadlines.

The key idea of improving performance is reducing the number of such half-finished jobs by using a grouping technique. The grouping technique decreases the number of half-finished jobs by allocating a job to multiple workers in a group at the same time. Although some adaptive grouping methods have been proposed for adaptive grouping, these methods are not applicable for credibility-based voting because they focus only on simple voting which cannot change the necessary number of results adaptively. In this chapter, an adaptive group-based job scheduling method is proposed for credibility-based voting with check-by-voting to improve the throughput of VC systems. The main contributions of this work are the follows. (1) Developing a dynamic grouping method, which predicts the optimal group size based on the expected-credibility and reduces the number of half-finished jobs. (2) Revealing the performance of VC systems with several job scheduling methods including the proposed method and m-first voting used in real VCs.

Chapter 7. Conclusions

This chapter concludes the results of this dissertation and discusses some promising research that can be further explored.

論文審査結果の要旨

近年，インターネット上の遊休計算資源を活用して大規模な並列計算システムを安価に構築するボランティアコンピューティング（VC）が注目されている．VC では，一部の参加者の妨害行為により計算結果の信頼性が著しく低下する可能性があるため，耐妨害性を備えることが必要不可欠である．著者は，信頼度に基づく多数決法に対し，信頼度の時間的変化に着目したジョブスケジューリングの高効率化手法と共に，抜取検査の不完全性を許容する実環境に即した信頼度計算法を考案し，その有用性を明らかにした．本論文はその成果を取りまとめたもので，全文７章よりなる．

第１章は，緒言である．

第２章では，対象とする VC システムの構成を述べ，その問題点を明らかにしている．

第３章では，信頼度に基づく多数決法の効率を改善する動的ジョブスケジューリング手法を提案している．本手法では，見込み信頼度という新たな指標を導入し，多数決に必要な冗長度を動的に推定することで，スケジューリングの高効率化を図っている．計算機シミュレーションにより，従来法と同程度の計算信頼性を保証しつつ計算時間の大幅な削減が可能なことを示しており，これは，計算の信頼性の保証と高性能化を同時に実現する有用な成果である．

第４章では，信頼度に基づく多数決法の抜取検査の頻度を最適化する高効率化手法を提案している．要求される計算信頼性に応じて，計算時間の期待値の理論的算出法と，それを最小化するための最適パラメータ導出法を示している．未知パラメータの存在を前提とした最適化方法論は，実際の VC システムの計算効率を大幅に向上できる重要な成果である．

第５章では，抜取検査の不完全性に着目した新たな信頼度計算法とそれを応用した多数決検査法を提案している．計算機シミュレーションにより，抜取検査の完全性を仮定した従来法の問題点を明らかにすると共に，提案手法が任意の未知パラメータに対して計算信頼性を保証可能なことを示している．これは実環境に即した仮定を導入可能にするものであり，実際の VC システムの信頼性保証を実現可能にする極めて重要な成果である．

第６章では，多数決検査の効率を改善するグループ化ジョブスケジューリング手法を提案している．見込み信頼度を用いて多数決を成立させる参加者グループを動的に構成することにより，従来法と同程度の信頼性を維持しつつ大幅に計算時間を削減可能なことを明らかにしている．本手法は，タスクの実行期限が定められている VC において，非常に有用な成果である．

第７章は，結言である．

以上要するに本論文は，耐妨害性を有する冗長計算を効率的に行うため，信頼度に基づく多数決の高効率化手法及び実環境に即した抜取検査の信頼度計算手法を考案し，その有用性を明らかにしたものであり，情報基礎科学の発展に寄与するところが少なくない．

よって，本論文は博士(情報科学)の学位論文として合格と認める．