IGI PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

ITJ 4916

This paper appears in the publication, International Journal of Grid and High Performance Computing, Volume 1, Issue 2
edited by **Emmanuel Udoh and Frank Zhigang Wang © 2009, IGI Global**

A Performance Study of Secure Data Mining on the Cell Processor

Hong Wang, Tohoku University, Japan
Hiroyuki Takizawa, Tohoku University, Japan
Hiroaki Kobayashi, Tohoku University, Japan

ABSTRACT

This article examines the potential of the Cell processor as a platform for secure data mining on the future volunteer computing systems. Volunteer computing platforms have the potential to provide massive computing power. However, privacy and security concerns prevent using volunteer computing for data mining of sensitive data. The Cell processor comes with hardware security features. The secure volunteer data mining can be achieved by using those hardware security features. In this article, we present a general security scheme for the volunteer computing, and a secure parallelized K-Means clustering algorithm for the Cell processor. We also evaluate the performance of the algorithm on the Cell secure system simulator. Evaluation results indicate that the proposed secure data clustering outperforms a non-secure clustering algorithm on the general purpose CPU, but incurs a huge performance overhead introduced by the decryption process of the Cell security features. Possible optimization for the secure K-Means clustering is discussed. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: Cell BE; Data Mining; Performance Evaluation; Secure Data Processing; Volunteer Computing

INTRODUCTION

The world's computing power is no longer primarily concentrated in supercomputer centers. Instead, it is distributed in hundreds of millions of personal computers and game consoles. The aim of volunteer computing is to use the Internet-connected individual computers to solve computing problems. Volunteer computing platforms have the potential to provide massive computing power. Currently, the most powerful volunteer computing platform - Folding@home (Folding@home Client Statistics, 2008) achieved more than one Petaflops computing power by connecting more than 600,000 PlayStation3. The second powerful volunteer computing platform - BOINC provided a sustained processing power of 350 Teraflops (Anderson, Christensen, and Allen, 2006) by November 2006. In contrast, the fastest conventional supercomputer, BlueGene/L achieves a maximum LINPACK performance of 478.2 Teraflops (The 30th TOP500 List, 2007). However, volunteer computing cannot be applied to the sensitive data processing, due to privacy and security concerns.

Latest generation game consoles are equipped with high performance processors. Therefore, these game consoles have the potential to become ideal peers of future volunteer computing systems. The CPU of PlayStation3 is the Cell processor, which was jointly developed by a Sony, Toshiba, and IBM alliance (Pham et al., 2005; Flachs et al., 2005). The Cell processor comes with hardware security features (Shimizu, Brokenshire, and Peyravian, 2005). The security features can be utilized to address the privacy and security concerns for sensitive data processing on the volunteer computing platforms.

Data mining is a collection of widely used techniques for extracting information from the enormous datasets. Nowadays, the increasing volume of data leads to the requirement for more computing power for data mining. Because data mining process can be well parallelized, it is an ideal application to take advantage of the volunteer computing. However, the data mining process on sensitive data has not been applied to the volunteer computing yet, due to the privacy and security concerns. On top of the hardware security features of the Cell processor, secure volunteer data mining can be achieved. In this article, we design a secure data processing method for volunteer computing on the Cell processor, and then apply it to an implementation of a typical data mining algorithm, called the K-Means clustering (MacQueen, 1967).

The rest of the article is organized as follows. Related work of secure data mining and our target application - K-Means clustering are introduced in Section 2. Section 3 introduces the Cell architecture, our previous work on performance evaluation of K-Means clustering on the Cell processor, and the hardware security features of the Cell processor. Section 4 presents a general security scheme for the volunteer computing, and a secure parallelized K-Means clustering algorithm for the Cell processor. Section 5 demonstrates a performance overhead that is introduced by security function, and discusses a solution for this performance issue. Section 6 concludes and summarizes the article.

RELATED WORK

Privacy preserving data mining (Verykios, Bertino, et al., 2004) is a novel research direction in data mining. The main object in privacy preserving data mining is to develop algorithms for modifying the original data in some way, so that the private data and private knowledge of any participants remain private even after the mining process. The typical modification methods include: perturbation, blocking, aggregation/merging, swapping, and sampling. A number of algorithms have been designed for different data mining techniques, such as classification, association rule discovery, and clustering. These algorithms can be classified into the following three types: