THE CHINESE UNIVERSITY OF HONG KONG
Department of Computer Science & Engineering and
Department of Information Engineering
Seminar

Sedic: Privacy-Aware Data Intensive Computing on Hybrid Clouds
by
Mr. Kehuan Zhang
Indiana University
USA

Date   :   **13 December 2011 (Tuesday)**

Time   :   **11:00am -12:00noon**

Venue :   **Room 833, Ho Sin Hang Engineering Building**

The Chinese University of Hong Kong

Abstract

The emergence of cost-effective cloud services offers organizations great opportunity to reduce their cost and increase productivity. This development, however, is hampered by privacy concerns: a significant amount of organizational computing workload at least partially involves sensitive data and therefore cannot be directly outsourced to the public cloud. The scale of these computing tasks also renders existing secure outsourcing techniques less applicable. A natural solution is to split a task, keeping the computation on the private data within an organization's private cloud while moving the rest to the public commercial cloud. However, this hybrid cloud computing is not supported by today's data-intensive computing frameworks, MapReduce in particular, which forces the users to manually split their computing tasks. In this talk, I will present a suite of new techniques that make such privacy-aware data-intensive computing possible. The system, called Sedic, leverages the special features of MapReduce to automatically partition a computing job according to the security levels of the data it works on, and arrange the computation across a hybrid cloud. Specifically, we modified MapReduce's distributed file system to strategically replicate data, moving sanitized data blocks to the public cloud. Over this data placement, map tasks are carefully scheduled to outsource as much workload to the public cloud as possible, given sensitive data always stay on the private cloud. To minimize inter-cloud communication, our approach also automatically analyzes and transforms the reduction structure of a submitted job to aggregate the map outcomes within the public cloud before sending the result back to the private cloud for the final reduction. This also allows the users to interact with our system in the same way they work with MapReduce, and directly run their legacy code in our framework. Sedic was implemented on Hadoop and evaluated using both real and synthesized computing jobs on a large-scale cloud test-bed. The study shows that our techniques effectively protect sensitive user data, offload a large amount of computation to the public cloud and also fully preserve the scalability of MapReduce.

Biography

Kehuan Zhang is a PhD candidate in the School of Informatics and Computing at the Indiana University Bloomington, USA. Prior to that, he worked as an FPGA logic R&D engineer for high performance routers and switches. He received BS and ME degrees in Computer Science from Hunan University, China in 2001 and 2004, respectively. His research interests are in the area of system, network, and software security, with a particular focus on security and privacy issues in desktop or web applications, mobile devices, cloud computing, and embedded systems.

**ALL ARE WELCOME **

Host: Professor Dah-Ming Chiu (Tel: 3943-8357, Email: dmchiu@ie.cuhk.edu.hk)
Enquiries: Information Engineering Dept., CUHK (Tel.: 3943-8385)