THE CHINESE UNIVERSITY OF HONG KONG
Department of Information Engineering

Seminar

Mining console logs
for problem detection in large scale systems

by

Mr. Wei Xu
Electrical Engineering and Computer Sciences Department
UC Berkeley

Date : **19 June, 2009 (Fri.)**
Time : **11:00am – 12:00noon**
Venue : **Room 833, Ho Sin Hang Engineering Building**
        The Chinese University of Hong Kong

Abstract
The console logs generated by an application contain information that the developers believed would be useful in debugging or monitoring the application. Despite the ubiquity and large size of these logs, they are rarely exploited because they are not readily machine-parsable.

We propose a novel approach for mining this source of information for system problem detection. We first combine log parsing and text mining with source code analysis to extract structure from the console logs. We then generate features from the structured information in order to detect anomalous patterns in the logs using a method based on Principal Component Analysis (PCA). Finally, we use a decision tree to distill the detection results to a format readily understandable by domain experts (e.g., developers, integrators and operators) who need not be familiar with the anomaly detection algorithms. The whole process requires no human intervention.

We evaluate our approach in two real-world systems, including Hadoop File System and Darkstar, an online game server, all of which produce accurate and easy to understand detection results.

In the talk, I will highlight the major points in console long mining as well as the new progress of doing the analysis on streaming logs.

Biography
Wei Xu is a PhD. candidate in EECS department of UC Berkeley. He works on the automatic log mining project in the RAD Lab. He is co-advised by Prof. David Patterson and Armando Fox. He has broad research interests in applying machine learning techniques to solve system problems, as well as distributed systems design. He did his undergraduate study in Tsinghua University and the University of Pennsylvania and master's at UC Berkeley.

** ALL ARE WELCOME **

Host: Professor Winston D.M. Chiu (Tel: 2609-8357, Email: dmchiu@ie.cuhk.edu.hk)
Enquiries: Information Engineering Dept., CUHK (Tel.: 2609-8385)