

THE CHINESE UNIVERSITY OF HONG KONG

Department of Statistics

will present a seminar entitled

**Modeling and Monitoring Data Arrivals in Distributed Data
Warehouses with Delays**

by

Professor Wei Jiang
Department of Industrial Engineering and Logistics Management
Hong Kong University of Science and Technology

on

Tuesday, 03 November 2009
2:00pm – 3:00pm

in

Lady Shaw Building G36
The Chinese University of Hong Kong

Abstract:

Data warehouse is a central repository of information collected from an array of distributed and heterogeneous operational database sources. However, there always exists delay between capturing operational data and making it available in a data warehouse. Poor data quality and the lack of integration in data warehouse make it difficult for business users to obtain a consistent view of business processes in areas such as finance, customer relationship management and supply chain operations. To monitor operational systems, this paper proposes a statistical model of data arrivals for data feeding systems from heterogeneous data collection stations to a central data warehouse. The statistical model is characterized by stochastic delays and irregular events caused by various random and business events. Due to the missing observations and aggregate data in the data feeding process, an augmented Kalman filter is proposed to track the data completeness and a statistical process control chart is developed to identify inconsistencies in data feeding systems. The effectiveness of the proposed method is demonstrated by an industrial example from a telecommunication data warehouse.

All are Welcome