

CALIFORNIA STATE UNIVERSITY, LONG BEACH

THE MATHEMATICS COLLOQUIUM

presents

Dr. Daniel Jeske

Statistics Dept., UC Riverside

speaking on

*Generation of Synthetic Data Sets for Evaluating the Accuracy of
Information Discovery and Analysis Systems*

Friday, April 28, 2006

12:00PM-1:00PM

FO3-200A

Abstract: Information Discovery and Analysis Systems (IDAS) are designed to correlate multiple sources of data and use data mining techniques to identify potentially significant events. Application domains for IDAS are numerous and include the emerging area of homeland security.

Developing test cases for an IDAS requires background data sets into which hypothetical future scenarios can be overlaid. The IDAS can then be measured in terms of false positive and false negative error rates. Obtaining the test data sets can be an obstacle due to both privacy issues and also the time and cost associated with collecting a diverse set of data sources.

In this talk, the design and development details of a prototype DAS Data Set Generator (IDSG) that enables a fast and comprehensive test of an IDAS will be reviewed. A person and credit card transaction application is used to illustrate the approach.

(This research was funded by a Department of Homeland Security Grant, No. HO32040450)