STATISTICS COLLOQUIUM

MONDAY, APRIL 19, 2010
TALK: 4:00 PM — SCIENCE CENTER, ROOM 309
RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

"Secure Statistical Analysis of Distributed Data"

Alan Karr
National Institute of Statistical Sciences

ABSTRACT

Secure multiparty computation protocols from computer science enable principled, secure analysis of horizontally partitioned, distributed databases when the sufficient statistics for the analysis are additive across database owners. When this structure is present, in fact only secure summation is required in order to perform a large number of analyses.

Following a discussion of secure analysis for vertically as well as horizontally partitioned data, the talk will focus on a number of unresolved--and in most cases, not even well-understood--issues relating to secure analysis of distributed data. These include data pre-processing, allowing analysts to be good statisticians, protections against dishonesty, risks arising from unequal database sizes and data heterogeneities, non-additivity, methods that allow agencies to opt out based on the results of the analysis, and numerical and algorithmic issues.