*The University of Alabama*
*Department of Computer Science*
*Colloquium Series*

**Dr. Brajendra Panda**
**Computer Science and Computer Engineering Department**
**University of Arkansas**

## Database Damage Assessment and Recovery following an Information Attack

**Monday, April 13th**
**11:00 a.m., EE 110**

### Abstract:

System invasion has become a common phenomenon especially after the concept of information sharing paved its way into the technical and business world. Databases have become one of the prime targets of attackers. The system under attack cannot distinguish an attacker from authorized users, and therefore would commit the attacker's transaction like any other valid transaction. Once damage is done, it would spread to unaffected parts of the database when valid transactions read damaged data and update clean data based on the value read. Therefore, it is necessary to develop faster and accurate Damage Assessment and Recovery (DAR) techniques to bring a damaged database back to consistent state as soon as possible.

Database log is the primary resource for DAR after an attack on the corresponding database. However, traditional logs do not store all operations of all transactions. DAR methods designed to restore databases from information attacks require that logs must store all operations of all transactions in order to identify the trail of damage and that the logs must not be purged. Therefore, the logs could be enormous in size and the DAR process would be slow due to the high volume of data to be accessed from the log.

This presentation will focus on various methods for damage assessment and recovery including log clustering, transaction dependency model vs. data dependency model, semantic logging, transaction fusion etc. Log clustering approach segments a log file based on transactions that have read-from relationships with other transactions in the cluster. Once a malicious transaction is identified, damage assessment mechanism accesses the corresponding cluster to determine the set of affected transactions. Recovery protocol then uses this list to maintain the consistency of the database. This eliminates the requirement of accessing operations of all transactions in the log, which may even be larger than the database. Log clustering can also be done using data dependency relationships that indicate how some data items are affected by other data items. Traditional logging protocols ignore many crucial operations and therefore accurate damage assessment is not possible using these logs. Semantic logging process eliminates this drawback by storing transaction semantics in the log. A malicious transaction may cause numerous valid transactions to be affected. In that case, during recovery, all these affected transactions must go through both UNDO and REDO procedures, which would slow down overall transaction execution process. By fusing multiple transactions together and executing them as one transaction improves system performance. These and several other DAR techniques will be discussed during the presentation.

### Bio:

**Dr. Brajendra Panda** received his M.S. degree in mathematics from Utkal University, India, in 1985 and Ph.D. degree in computer science from North Dakota State University in 1994. After completion of his Ph.D., he joined the Computer Science Department of Alabama A&M University and in 1997 he moved to the University of North Dakota. Then in 2001, he joined the Computer Science and Computer Engineering Department of the University of Arkansas, where he is currently a full professor. His primary research interest is database security although he has worked in other related areas such as information provenance, trusted computing etc. He has published over 90 research papers in these areas and has obtained over $2.4 million in research funding as a principal investigator. He has also served as program committee member on numerous conferences and workshops.