# Fault Detection in Large-Scale Distributed Systems

## by Geoff Jiang, NEC Laboratories America, Princeton, NJ

**Date:**      September 21, 2005 (Wednesday)
**Time:**      4:45 pm (refreshment starts at 4:30 pm)
**Place:**      202 ECEC, NJIT

### About the Speaker

Dr. Geoff (Guofei) Jiang is a research staff member with the Robust and Secure System Group at NEC Laboratories America, Princeton. Before Joining NEC Labs in 2004, he was a senior research scientist in the Institute for Security Technology Studies at Dartmouth College and worked on several multi-million dollar projects funded by DARPA, DHS and ARDA. He got his BS and PhD in ECE from Beijing Institute of Technology and was a Postdoctoral Fellow in Computing Engineering at Dartmouth College. His research focus is on large-scale distributed system, dependable and secure computing, system and information theory. He has published over 40 papers and has several US patents pending.

### About the Talk

The increasing complexity of today's systems makes fast and accurate failure detection essential for their use in mission-critical applications. Various instrumentation methods provide a wealth of information that can be used to model the system's normal behavior. Any deviation from this behavior may be indicative of failure. In this talk the lecturer will present two complementary techniques for anomaly detection. The first is based on modeling the request traces through the system. Varied-length n-grams and automata are used to characterize the normal traces. Training data is used for automatically extract automata with various resolutions. A new trace is compared against the learned automata to determine whether it is abnormal. The second is based on tracking over time the frequency of the interaction between any two components in the system. We decompose the observation data into signal and noise subspaces and use two statistics, the Hotelling T square score and squared prediction error (SPE) to represent their character-istics. Instead of tracking the original data, we use sequentially discounting expectation maximization (SDEM) algorithm to learn the distribution of the extracted statistics. A failure event can then be detected based on the abnormal change of the distribution. Both approaches have been tested in a real system with injected faults and achieved good results in fault detection experiments.

**Sponsors:**      **IEEE Signal Processing Society North Jersey Chapter**
      **IEEE Communications Society North Jersey Chapter**
      **NJIT Department of Electrical and Computer Engineering**

For more information contact Yun Shi (973)-596-3501, Nirwan Ansari (973) 596-3670, Alfredo Tan (201) 692-2347, and Hong Man (201)-216-5038,  or check  **http://web.njit.edu/~ieeenj/comm.html** for latest updates. Directions to NJIT can be found at: **http://www.njit.edu/University/Directions.html**.