ELECTRICAL & COMPUTER ENGINEERING BIO-SEMINAR Spring 2017 | ENGINEERING TEXAS A&M UNIVERSITY

When: **Friday 15:00 – 16:00**

Where: **ETB 1003**

Speaker: **Prof. Krishna Narayanan**

Eric D. Rubin '06 Professor
Department of Electrical & Computer Engineering
Texas A&M University

Title: **Pattern Matching via Sparse Fourier Transforms -- A Low Complexity Algorithm for Big Data**

Date: **04-07-2017**

Abstract: We are witnessing an unprecedented growth in the amount of data that is being collected and made available for data mining. While the availability of large-scale data sets presents exciting opportunities for advancing sciences, health care, understanding of human behavior etc., mining the data set for useful information becomes a computationally challenging task. We are in an era where the volume of data is growing faster than the rate at which available computing power is growing, thereby creating a dire need for computationally efficient algorithms for data mining. From an algorithmic complexity stand point, we are transitioning from a mindset where algorithms with linear complexity in the size of the data set were considered efficient to an era when algorithms with linear complexity have become infeasible owing to the large size of the data sets. This necessitates the creation of algorithms with sub-linear time complexity tailored to big data.

One of the most fundamental data analytics tasks is that of querying a data set to see if a particular pattern of symbols appears in the data set either exactly or approximately. We assume that sketches of the original signal can be computed off line and stored. Using the sparse Fourier transform computation based approach introduced by Pawar and Ramchandran, we show that all such matches can be determined in sub-linear time with high probability (asymptotically in the size of the data set and the query).

I will provide a brief overview of the results in fast pattern matching before presenting our work. I will try to keep the main ideas in the talk fairly accessible to any one with an undergraduate-level exposure to Fourier transforms and I will try to keep the talk fairly self-contained. Potential applications of this work include text matching, audio/image matching, DNA matching in genomics, metabolomics, radio astronomy, searching for signatures of events within large databases, detecting viruses within binary executable files. I am actively looking for collaborators who can use fast pattern matching in their area of expertise.

Biography: Krishna Narayanan is a professor in the ECE department at Texas A&M University. His research interests are in coding theory, information theory, and signal processing with applications to wireless communications, data storage and data science. On the teaching side, he is excited by the use of technological tools to personalize the learning experience of students. When he is not pattern matching at work, he (mostly unsuccessfully) matches patterns to identify ragas when listening to Indian classical music. He is also a self-proclaimed expert in analyzing cricket matches.