Abstract:
Efficient end-to-end parallel/distributed processing of vector-based spatial data has been a long-standing research question in GIS community. This has not been for the lack of individual parallel algorithms, but as we discovered, it is because of the irregular and data intensive nature of the underlying computation. While effective systems for parallel processing of raster-based spatial data files are abundant in the literature, there is only meager amount of reported system work that deals with the complexities of vector (polygonal) data and none on cloud platform. We have created an open-architecture-based system named Crayons for Azure cloud platform using state-of-the-art techniques. Cloud platform is well suited for GIS scientists due to web-based accessibility and on-demand scalability. The design and development of Crayons system is an engineering feat both due to (i) the emerging nature of the Azure cloud platform which lacks traditional support for parallel processing and (ii) the tedious exploration of design space for right techniques for parallelizing various workflow components including file I/O, partitioning, task creation, and load balancing. We present detailed architectural designs of Crayons system exploring three static and dynamic task creation, allocation and load balancing sub-components. We demonstrate how Azure platform's computation, communication, and storage mechanisms can be utilized for scientific high performance computing (HPC) applications. Crayons scales well for sufficiently large data sets, achieving end-to-end speedup of over 27-fold employing 100 Azure processors (101K polygons in base layer intersected with 128K polygons in overlay layer in two GML files). For smaller, more irregular workload, it still yields about 10-fold speedup (4K and 465K polygons). I will also point to open problems, including parallelization of R-Tree data structure on GPGPUs.

About the Speaker:
Sushil K. Prasad (BTech’85 IIT Kharagpur, MS’86 Washington State, Pullman; PhD’90 Central Florida, Orlando - all in Computer Science/Engineering) is a Professor of Computer Science at Georgia State University (GSU) and Director of GSU-GEDC Distributed and Mobile Systems (DiMoS) Lab hosted at Georgia Institute of Technology, Atlanta. He has carried out theoretical as well as experimental research in parallel and distributed computing, resulting in 80+ refereed publications, several patent applications, and about $1M in external research funds as PI and over $4M overall (NSF/NIH/GRA/Industry). Recently, Sushil successfully led a multi-year, Georgia Research Alliance (GRA) funded interdisciplinary research project with seven GSU faculty, three Georgia Tech faculty, and over two dozen students on developing SyD middleware for collaborative distributed computing over heterogeneous mobile devices, resulting in several patents, dissertations, and publications. Sushil has been very active in the professional community, serving on the organization of top conferences, on NSF and other review panels, on advisory committees of conferences and State of Georgia funding agency Yamacraw, and carrying out editorial activities of conference proceedings and journal special issues.

All are welcome! Refreshments are provided.
For enquiries, please call 2859 2180 or email enquiry@cs.hku.hk
Department of Computer Science
The University of Hong Kong