

Seminar: Friday 9 August 2013: Monash Undergraduate Research Projects Abroad (MURPA).

The annual MURPA program exposes final year undergraduate students to an international research experience within a leading research laboratory. This is the first year University of Warwick has taken a Monash student.

On Friday 2 August Dr Sameer Tilak will present his seminar "Storage and Analysis of Big Data from Sensor Networks: Challenges and Opportunities" live to audiences in Melbourne and Brisbane.

Melbourne Date: Friday 9 August Time: 9.00 - 10.00am Location - Monash University: Seminar Room G12A, Building 26, Clayton Campus Enquiries: Caitlin Slattery (Faculty of IT)

Brisbane Date: Friday 9 August Time: 9.00 - 10.00am Location - University of Queensland: Seminar Room 505A, Building 47 (Axon), St Lucia Campus Enquiries: Sharon Cook (Research Computing Centre)

San Diego Date: Thursday 8 August Time: 4.00pm - 5.00pm Location: 5004 Meeting Room Enquiries: Teri Simas

Abstract: Storage and Analysis of Big Data from Sensor Networks: Challenges and Opportunities

Sameer Tilak

Large-scale environmental observing systems consist of sensors embedded deeply with our physical environment. Typically these systems consist of highly-calibrated sensors deployed at strategic locations to generate science quality data. Recently, personalized mobile sensing (e.g., sensors mounted on cars or carried by people) has received considerable attention from the research community. Although, these systems consist of cheap, mobile sensors that are not well calibrated, they can provide spatial sampling diversity as compared to the traditional environmental observing systems. Together these large-scale sensor networks can gathering data at high spatio-temporal resolutions and have potential to provide scientists unprecedented insights into complex physical environment. However, managing and processing of this "big data" presents challenges and research opportunities. To that end, Solid-state (Flash) disks present an attractive storage option. Solid-state (Flash) drives are becoming cheaper and more common in Data Centers and we believe that this trend will continue to grow. By 2020, the quantity of electronically stored data will reach 35 trillion gigabytes. Big data technologies such as Apache Hadoop, HBase, Pig, Hive are striving to make the storage, manipulation and analysis of huge volumes of data cheaper and faster than ever. With current 6GBps SATA III, NAND-based Solid-State (flash) drives are delivering astounding performance (in terms of random I/O) compared to traditional hard-disk drives. However, their benefit for big data processing is not yet quantified.

Goal of this research is to understand how the Big data technologies can benefit from the emerging technology, namely, solid state drives in the context of storage and analysis of data gathered by sensor networks. In particular, we will explore tools such as Apache Hadoop,

Fig, Cassandra for processing and storage of data for large-scale sensor network applications. We will use FutureGrid (a national-scale testbed) to conduct the benchmarking experiments.