**COMPUTER SCIENCE & ENGINEERING**

[**HOME**](../CONTENTS.htm)

Author : **SAI KRISHNA**

Title of the thesis : **ENERGY EFFICIENT AND RELIABLE ALGORITHMS FOR DATA GATHERING IN WIRELESS SENSOR AND ACTOR NETWORKS**

Guide : **Dr. RASHMI RANJAN ROUT**

Degree : **Ph. D.**

Student ID No. : **701431**

**ABSTRACT**

Wireless Sensor and Actor Networks (WSANs) recently emerged as major information gathering paradigm due to its wide variety of applications, such as issuing tsunami alerts, chemical attack detection, forests fire detection and intrusion detection in military surveillance. The energy-constrained sensor nodes spend more energy in transmitting data packets than in sensing operation. The data transmission rate and delivery delay increase due to increase in packet dropping rate. Thus, improvement in energy efficiency is a challenging issue while providing reliable data delivery and stringent delivery delay in WSANs. Reduction of packet dropping rate results in improvement of energy efficiency, data transmission reliability and delivery delay. Packet dropping occurs mainly due to unavailability of free buffer and unreliable wireless links. Further, energy harvesting technology extends the lifetime of a sensor network. The volume of harvested energy varies dynamically with the change in weather conditions over time. This may lead to temporary disconnection improves the reliable data transmission rate in energy harvesting sensor actor networks. The thesis focuses on the energy efficient and reliable algorithms for data gathering in wireless sensor and actor networks. The issues of energy efficiency and reliable data transmission have been addressed while maintaining stringent data delivery delay. In this thesis, the proposed approaches have achieved reliable data transmission by reducing the packet dropping rate. Firstly, a mark decision process based buffer management mechanism has been designed in tree-based WSAN to reduce the data delivery delay and to improve the energy efficiency.

 A mathematical model for a mobile actor is presented to analyze buffer occupancy and energy consumption with event-centric traffic. Secondly, a reliable data transmission mechanism using opportunistic encoding has been proposed for a WSAN with faulty nodes. The proposed mechanism analyzes the quality of link states and determines the applicability of network coding to improve the data transmission reliability and to reduce the number of data transmissions. Thirdly, a fuzzy based delay and energy-aware intelligent routing mechanism has been developed to take effective routing decisions. The proposed mechanism computes a routing metric by considering the residual energy, link quality, available buffer and distance. Finally, a fuzzy based adaptive duty cycling algorithm has been designed to achieve network sustainability in harvesting sensor actor networks. Energy consumptions of bottleneck zone has been estimated with the proposed network model.

 The proposed mechanism takes switching decision for network survivability. Coordinated duty cycle schedule is considered to improve the reliable data delivery. Performances of proposed approaches have been evaluated through simulation.

**Keywords:**Wireless sensor and actor networks, energy efficiency, delay, reliable data transmission, buffer management, markov decision process, network coding, energy and delay aware routing, energy harvesting nodes, network sustainability, duty cycle, fuzzy logic.

 **COMPUTER SCIENCE & ENGINEERING**

Author : **SANDIPAN MAITI**

Title of the thesis : **MAP-REDUCE BASED SPATIAL DATA MINING SYSTEM FOR CO- LOCATION AND BEHAVIOURAL PATTERNS**

Guide : **Dr. R.B.V. SUBRAMANYAM**

Degree : **Ph. D.**

Student ID No. : **701367**

**ABSTRACT**

Recent advancement in science and technology has made human life better and more comfortable through developing various important and required applications. Innovation of GPS technology and their integration within mobile devices and sensor is the source of huge amount of data, which is streamed towards data centres, applications for further processing and decision making, supported by advances networking technologies. For example mobile phones, digital cameras, automobiles etc. are equipped with GPS, various types of sensors and wireless networking module. GPS is capturing geographical location of specific objects, various sensors are capturing different features parameters like temperature, pollution etc. and transmits these data to nearest data centres using wireless networking module. This collection is named as spatial data, as it will have data values towards a define feature set along with geographic location of these spatial objects with timestamp.

 Spatial data is mined for extraction of patterns, like Co-location patterns and Behavioural patterns, are set of entities often found within close proximity or expressing similar behaviour with respect to define time window. These pattern mining algorithms need to redesign for processing this distributed Spatial Data, which is being characterised as Big Data according to Big Data VVV definition and HACE theorem. The research works presented in this thesis has motivation towards proposing a Data Mining Subsystem as part of recommendation system, and efficient distributed algorithms for processing of Big Spatial Data for mining Co-location, Behavioural patterns.

 Firstly, a Map Reduce based distributed algorithmic framework has been proposed for mining Co-location patters by considering locations of spatial objects from Big Spatial Data. A temporal mining algorithm is also proposed for capturing the changes in patterns and increase the mining efficiency, when newly collected data is being added or the time window has moved forward. The incorporation of temporal pattern mining algorithm will reduce the cost of data processing in this framework. Secondly, a Behavioural Similarity measure has been proposed to measure the similarity among spatial objects by considering the set of feature values in Spatial Dataset. A Behavioural pattern mining algorithm is also proposed here using Behavioural Similarity measure.

 Thirdly, a density based clustering algorithm has been proposed for distributed spatial data using Kernel Density Estimator. This clustering algorithm is able to produce data grids from streaming spatial data, which will improve the mining efficiency of Co-location and Behavioural pattern mining algorithms. Finally, it is aimed to propose a Data Mining sub-system for mining Co-location and Behavioural patterns from data grids, generated by KDE based distributed clustering algorithm from distributed spatial data. These mined patterns will be used for making meaningful and important recommendations for the end-users.

 **COMPUTER SCIENCE & ENGINEERING**

Author : **SUMALATHA SALETI**

Title of the thesis : **MINING BIG DATA FOR FREQUENT PATTERNS USING MAPREDUCE COMPUTING**

Guide : **Dr. R. B. V. SUBRAMANYAM**

Degree : **Ph. D.**

Student ID No. : **715069**

**ABSTRACT**

 The main motivation of frequent pattern mining is to extract useful patterns from the data sets. Interesting associations among the data can be discovered by mining the frequent patterns. Among the different kinds of pattern mining, frequent item set mining has been applied widely in many applications such as market basket analysis, medical applications, online transactions, social network analysis and so forth. An item set is called frequent if the set of items in it appear frequently together. However, frequent item set mining can find only the frequent item sets, the time regularity of the items cannot be found. Sequential pattern mining considers both the frequency of the items and the order of items based on their time stamps. It attracted great deal of attention in many applications such as customer buying trend analysis, web access mining, natural disaster analysis and so forth.

 The patterns mined from sequential pattern mining algorithms do not consider the cost or profit of the item. A sequence that is not frequent in a dataset may contribute much to the overall profit of the organization due to its high profit. Hence, utility sequential pattern mining considers quantity and timestamp of items as well as profit of each item. Because of constantly arriving new data, the resultant patterns of frequent pattern mining may become obsolete over time. Hence, it is necessary to incrementally process the data in order to refresh the mining results without mining from scratch. The advancement in technology led to the generation of huge volumes of data from multiple sources such as social media, online transactions, internet applications and so forth.

 This era of big data pose a challenge to explore large volumes of data and extract the knowledge in the form of useful patterns. Moreover, the conventional methods used in mining patterns are not suitable for handling the big data. Hence, in this thesis, we investigate the solutions for frequent pattern mining on big data using a popular programming model known as Map Reduce. Firstly, we propose a parallel algorithm for compressing the transactional data that makes the data simple and Bit Vector Product algorithm is proposed to mine the frequent item sets from the compressed data.

 Secondly, distributed algorithm for mining sequential patterns using conccurrence information is proposed. Here, we make use of item co-occurrence information and reduce the search space using the pruning strategies. Thirdly, distributed high utility time interval sequential patterns with time information between the successive items are mined. Finally, an incremental algorithm is proposed to make use of the knowledge obtained in previous mining while mining sequential patterns. All the proposed algorithms are tested on our in house Hadoop cluster composed of one master node and eight data nodes.

**Keywords:** Big Data, Data Mining, Frequent Pattern Mining, Frequent Item set Mining, High Utility Sequential Pattern Mining, Incremental Mining, Map Reduce, Sequential Pattern Mining.