

## CHAPTER 12

### COMPUTER SCIENCE

#### Doctoral Theses

01. BADHANI (Shikha) nee Shikha Malik  
**Techniques for Android Malware Detection**  
Supervisor : Prof. Sunil Kumar Muttoo  
Th 24562

*Abstract*  
(Not Verified)

Android has been at the dominant position consistently for the past few years and consequently, its prevalence has also led to an escalation of malware in Android applications. The primary goal of this thesis is to develop techniques for creating robust Android malware detection systems that can detect malwares efficiently and correctly. To analyze the level of security provided by the popular commercially available anti-malwares, various techniques of hiding a known malware inside an image are presented. Popular anti-malwares are then used to scan the malicious variants created using hiding techniques. Most of the anti-malwares failed to detect even trivial techniques used to hide the malware in an image. This analysis propelled us towards creating Android malware detection systems that can detect such stealthy malwares. In the experiments presented in this thesis, performing ensemble learning post-classification is found more advantageous than pre-classification since the post-classification ensemble is immune to the threshold used for creating subsets of features. An Android malware detection system – CENdroid is presented which combines clustering along with ensemble learning to achieve better performance metrics as compared to base classifiers as well as their ensembles. To capture the semantics of the code, an algorithm of extracting graph-based features from Android applications – code graphs, is presented in this thesis. Ensemble learning has been further extended to graph-based base classifiers (using code graphs as features) and an Android malware detection system – GENdroid is presented that combines various graph-based base classifiers to achieve better performance as compared to individual graph-based classifiers. A framework is presented to evaluate the proposed systems – CENdroid and GENdroid against various obfuscation tools as well as the techniques presented to hide malware inside an image. Both CENdroid and GENdroid displayed resilience against obfuscation and app hiding techniques at both, feature-level as well as classifier-level.

#### *Contents*

1. Introduction 2. Android malware 3. Evading Android anti-malware by hiding malicious application inside images 4. Comparative analysis of pre- and post-classification ensemble methods for android malware detection 5. CENroid—cluster-ENsemble classifier for detecting malicious android applications 6. Android malware detection using code graphs 7. GENroid—graph-based ENsemble classifier for detecting malicious android applications 8. Analyzing CENdroid and GENdroid against code obfuscation and app hiding techniques 9. Conclusion and future work, Appendices. Bibliography.

02. CHARTTERJEE (Indranath)  
**Automated Diagnosis of Schizophrenia and Identification of Affected Brain Regions**

Supervisor : Prof. Naveen Kumar and Dr. Manoj Agarwal

Th 24563

*Abstract  
(Verified)*

Schizophrenia is a serious mental disorder that manifests functional and biochemical changes in the brain. Neuroimaging techniques such as magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI) play important role in the diagnosis of schizophrenia. The main challenge in fMRI/MRI data analysis lies in the high dimensionality of the data. In this thesis, we have made an effort to identify the brain regions affected by schizophrenia and developed computer-aided tools to distinguish between schizophrenia and healthy subjects. We have proposed a novel three-stage approach for identification of affected brain regions that distinguish between schizophrenia and healthy subjects. Beginning with the application of general linear model (GLM), we apply statistical hypothesis testing, and deploy bi-objective optimization approach. The proposed approach serves as the basis for developing a computer-aided diagnosis tool for classifying the schizophrenia and healthy subjects with high accuracy. The study is able to identify certain brain regions showing the subtle difference in functional activations between the young and old schizophrenic patients. The study also reveals some brain regions not reported earlier. We also study the working memory dysfunction in schizophrenia with a view to identify the affected brain regions using group independent component analysis (GICA) that achieve high classification accuracy. Traditional computation-time intensive methods such as GLM and ICA are often used to transform 4-D time series fMRI data to 3-D spatial brain maps. We have developed a mean-deviation based approach that yields high classification accuracy. Finally, we apply standard voxel-based morphometry to identify the grey matter changes in conjunction with NSGA-II to select a set of voxels that show significant atrophy in grey matter concentration in several brain regions with high classification accuracy.

*Contents*

1. Introduction 2. Bi-objective approach for computer-aided diagnosis of schizophrenia patients using fMRI data 3. Impact of ageing on the brain regions of the schizophrenia patients: An fMRI study evolutionary approach 4. Identification of brain regions associated with working memory deficit in schizophrenia 5. Meandeviation based identification of activated voxels from time-series fMRI data of schizophrenia patients 6. Identification of changes in grey matter volume using an evolutionary approach: An MRI study of schizophrenia 7. Conclusions and future works, Bibliography.

03. GUPTA (Ranjan)  
**Techniques for improvement of E-Governance in Developing Nations**

Supervisors : Prof. Sunil Kumar Muttoo and Saibal Kumar Pal

Th 24564

*Abstract  
(Verified)*

Engagement to ensure the full utilization of E-Governance. The current research work is dedicated towards addressing some of these concerns through four research objectives, specifically for developing nations.

The first objective addresses the lack of region specific development assessment model for a developing nation, where the setup is different as compared to developed nations. The second objective is related to E-Governance related Infrastructure management and Facility Location Allocation problem is solved through combination of various meta-heuristic algorithms. The third objective addresses various internal and external security concerns for E-Governance setup in developing nation. An internal security framework is designed for organization and an Intrusion detection system is developed using meta-heuristic algorithms for better monitoring of external security. The fourth objective is related to maximization of citizen engagement with government through better electronic participation process. A better service utility model is proposed from citizen's point of view and efficient usage of social media & data analytics is proposed under the last objective of this study. The datasets have been collected from official websites of Indian Government for testing of various techniques proposed in different chapters. Data points were also mentioned for other developing nation like South Africa. All the experiment work has been carried out using technologies like MATLAB, R, LINGO, and WAMP, on a DELL Inspiron 15R Laptop with 2 GB RAM and 500 GB Hard Disk, and having Intel i5 1.70GHz processor. This work would be useful for researchers from community of Computer Science, Data Science, Information Technology, Information Systems, and Polity/Governance.

#### *Contents*

1. Introduction 2. Literature review 3. Components and index for regional development assessment of E-governance 4. Location allocation of E-governance infrastructure 5. Internal security of E-governance systems 6. External security and monitoring for E-governance systems 7. Improving E-participation of citizens 8. Discussion & conclusion, Bibliography, Appendices.

04. HIMANI

#### **Securing digital Information using Crypto-stego Methods**

Supervisor :Prof. Naveen kumar and Prof. R.K. Sharma

Th 24560

#### *Abstract (Not Verified)*

With the rapidly increasing usage of internet, the need for cyber security concerns has also increased. Towards this end, we have proposed various crypto-stego systems. We have proposed symmetric key based crypto-stego systems that use adaptive Vigenere and one time pad Vigenere approaches. Encrypted data is transformed using FFT to provide protection from partial data attacks. Resultant values are hidden in an image using spatial steganographic technique using LSB-M. Statistical analysis proves the strength of the proposed approach. We have also proposed a crypto-stego system with variable length key encryption using modified affine transform approach to secure information. Data is partitioned and encrypted with different keys in different parts. The proposed approach provides a large key space. To prevent partial data attacks, DCT has been used. The proposed system produces high quality images, without any data loss. We have proposed another high capacity steganographic technique for images. DWT converts image spatial values to spectral coefficients. These values are compressed and hidden in the scrambled image produced by Arnold transform. Encrypted data is spread all over the image and appears as noise to an attacker. Although, this technique is lossy, yet it yields a high-quality image. We have used Rabin cryptosystem to encrypt information and the encrypted information is hidden in frequency coefficients using LSB-M technique obtained by integer wavelet transform. The results show that this approach yields minimal distortion in cover

image and is immune to various types of steganalysis attacks. We have also proposed a multi-image steganography and authentication approach using crypto-stego techniques. The secret information encrypted using Rabin is randomized using Arnold transform and subsequently distributed in multiple images using minimal key and distribution key using edge adaptive technique. The proposed crypto-stego methods have been shown to be strong against various steganalysis and cryptanalysis attacks.

*Contents*

1. Introduction 2. Crypto-stego system for security of two dimensional data 3. Securing digital information using variable length key in crypto-stego system 4. Hiding a color image in scrambled image 5. Securing digital information with minimal distortion in an image 6. Multi-image steganography and authentication using crypto-stego techniques 7. Conclusion, Bibliography.

05. JOHARI (Rahul)

**Routing Schemes for Delay Tolerant Networks**

Supervisor : Prof. Neelima Gupta and Sandhya Aneja  
Th 24559

*Abstract*  
(Not Verified)

Disruption and Delay Tolerant Network(DTN) is a networking architecture that was designed to provide connectivity, similar to Internet, in challenging environments characterized by long delays, partitioned / sparse networks and intermittent connectivity. Routing a message in networks that are dynamic in nature with time varying partially connected topology has been a challenge. The heterogeneity of the types of contacts available in such networks also adds to the complexity. Most of the existing approaches exploit either the opportunistic contacts and transfer messages using the probabilities of delivering a message or use periodic contacts. It is understood that considerable improvement in message delivery can be achieved by using more than one type of contact and/or by making better estimate of node's delivery probability in a given scenario. This thesis work presents three novel routing schemes that uses more than one type of contact and multiple nodes' attributes to improve the routing performance in DTNs. It was further observed that almost all the existing routing schemes have been evaluated on three basic metrics which seem to be quite deficient in evaluating them. To have a better understanding of the performance of routing schemes it is necessary to evaluate them on other aspects such as resource utilization, scalability, security etc. To overcome this deficiency, a metrics set has been proposed to evaluate the routing schemes. The proposed metrics measures resource utilization, scalability, performance and security.

*Contents*

1. Introduction 2. Disruption and delay tolerant network 3. Routing in DTN 4. Dynamic social grouping based routing with periodic carriers 5. CONCOR-Context aware community oriented routing 6. POSOP: A DTN routing scheme 7. Experimental evaluation of routing schemes 8. Conclusions and future work, Bibliography.

06. SACHIN KUMAR  
**Modelling and Prediction of Building Energy Performance using Machine Learning Approaches.**  
 Supervisor :Dr. Ram Pal Singh  
Th 24561

*Abstract*  
*(Not Verified)*

Amount of energy which needs to be optimized, monitored, estimated and predicted and properly utilized in order to improve the efficiency and comfort of buildings. Energy performance monitoring and maintenance depend on parameters such as ambient environment, outside environment, equipment settings and usage patterns, and occupancy and activities. An accurate modelling, prediction and monitoring of the building's energy performance can help in alerting the maintenance staff when it identifies some wasteful, unreliable, faulty and hazardous trends. This helps in solving the problems and taking actions before the issues become big and cost more. Energy consumption and performance prediction is a complex sophisticated problem that needs advanced, adaptive and efficient prediction and modelling techniques. Previously developed, Mathematical and equation-based models face difficulties as they are rigid, not very adaptive to the dynamic nature of constantly changing parameters. The real-time dynamics of domain demand that the prediction and modelling should be highly accurate and efficient, generalised, online and adaptive. This thesis presents novel methods based on modern machine learning architecture extreme learning machine(ELM) and its variants. Traditionally, Artificial Neural Networks(ANNs) models have been applied by many researchers but they suffer from their variability in SLFNs for different circumstances, manual parameter tuning, over-fitting and generalisation, local minima and time-consuming nature and computationally inefficiency. Proposed models are developed on the shallow architecture, take into consideration variability of parameters, dynamicity of environment, online/real-time data processing and learning, fast and efficient computation.

*Contents*

1. Introduction 2. Background and state of art 3. Heating and cooling load prediction through building design and structural attributes 4. Ensemble model to predict energy performance of buildings 5. Modelling and prediction of energy consumption and occupancy detection using sensors 6. Hybrid model for short-term temperature prediction 7. Conclusion and future work, References.

07. SAXENA (Rakhi)  
**Hierarchical Graph Decomposition for Scalable Network Analysis**  
 Supervisor :Dr. Sharanjit Kaur and Prof. Vasudha Bhatnagar  
Th 24558

*Abstract*  
*(Not Verified)*

Revealed by the decomposition mimic community membership. These topological signals are leveraged to develop solutions for three important problems in network analysis, namely, network comparison, computation of node centrality and detection of influential spreaders. The task of network comparison entails quantification of distance between network signatures. We propose algorithm

Network Comparison via k-core Decomposition (NCKD) that uses probability distribution of nodes and edges at levels of hierarchy to compose network signature. We further refine NCKD algorithm by using node-level assortativity and more expressive methods of distribution aggregation to propose algorithm Network Similarity using Hierarchical Decomposition (NSD). Since humans derive benefits concomitant with their position in the network hierarchy, and with the strength of their connections within and between communities, we propose Social Centrality (SC) score that emulates the real-life behavior of social actors to bond within community and bridge between communities. Method k-truss decomposition is applied to elicit network hierarchy and discriminate between intra- and inter-community ties. We also propose a measure named Influence score using Position, Reachability, and Interaction (IPRI) that incorporates position of the actor in the network hierarchy, intensity of his interactions with neighbors and extent of actor's connectivity in different levels of hierarchy. It is concluded that in the era of massive networks, hierarchical graph decomposition is a versatile approach for social network analysis.

#### *Contents*

1. Introduction 2. Preliminaries 3. Hierarchy-based network comparison 4. Improved signatures for network comparison 5. Social centrality 6. Identification of influential spreaders 7. Concluding remarks.