JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Enhancing the Performance of a Wireless Network through Advanced Traffic Analysis

Indira Umarji[1], Dr. S M Joshi[2], Soumya S Katageri[3], Amulya U Naik[4], Rashmi S S[5], Bhuvan B[6]
[1]Asst. Prof., [2]Professor, [3, 4, 5, 6] B.E student
Computer Science and Engineering
SDM College of Engineering and Technology, Dharwad, India

Abstract: In the realm of wireless network management, the selection of appropriate algorithms plays a crucial role in enhancing performance metrics such as throughput, and security. This research project delves into a comprehensive study of various network algorithms, including K-means clustering, Markov Models, Time Series Analysis, Random Forest, and Fourier Series, Hidden Markov Model, Random Walk to evaluate their efficiency in optimizing network performance across varying numbers of nodes. The methodology involves implementing each algorithm meticulously and even evaluating its performance under diverse network configurations. Scrutinizing algorithm outputs throughput and drop ratio reveals trends, strengths, and weaknesses. Results are synthesized graphically, enabling a comparative assessment of algorithm performance across different network scenarios. This study provides valuable insights into algorithm selection for wireless network optimization. After the execution of algorithms with multiple iterations the average drop ratio is 9.55 percent and average throughput is 90.055 percent. By elucidating algorithm capabilities and limitations, this research contributes to improved real-world network management strategies.

Keywords:
Network algorithms ,throughput, drop ratio, performance evaluation.

I. INTRODUCTION

Due to huge number of applications significance of Wireless Sensor Networks (WSNs) in various fields like agriculture, security, and automation because of their ability to monitor environmental conditions and communicate wirelessly.Each node in a WSN is responsible for delivering data to a destination node, but this process consumes a significant amount of energy, leading to node failures over time.

The development of low-cost sensor networks has expanded their applications, and researchers are continuously addressing technical challenges. These networks, composed of small, multifunctional sensors, can collect, process, and distribute information, with routing algorithms playing a key role in data gathering and aggregation. This shows the importance of guaranteed data delivery in Wireless Sensor Networks (WSNs) and the significant energy consumption associated with data transmission. It emphasizes the challenge of node failure due to energy depletion and suggests that routing data efficiently can mitigate this issue.

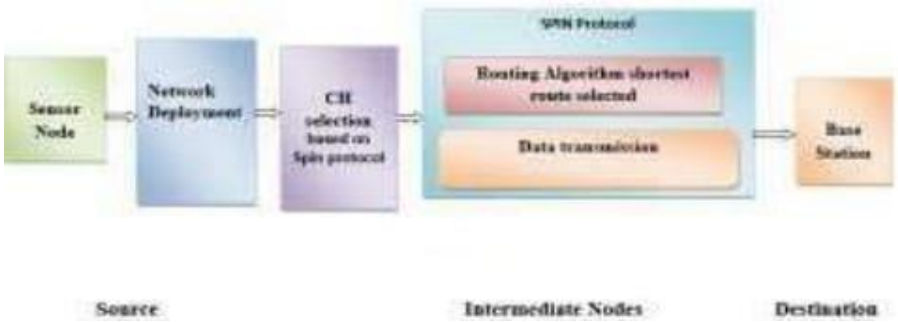

Figure 1.1 System Architecture[6]

In the rapidly evolving landscape of wireless communication, the project, "Enhancing the Performance of a Wireless Network through Advanced Traffic Analysis," emerges as a pivotal endeavor. With the escalating demand for high-speed and reliable connectivity, the challenges faced by network administrators in optimizing key performance parameters such as drop ratio, throughput, and other metrics are more pronounced than ever. This project aims to bridge this gap by developing a sophisticated Wireless Network Traffic Analyzer. The multifaceted approach includes elements such as advanced packet sniffing, traffic classification, and the integration of machine learning algorithms. By delving into the intricacies of packet-level analysis, the analyzer not only identifies congestion points but also classifies traffic types, enabling targeted optimization. Through real-time monitoring and proactive optimization, the project not only seeks to reduce drop ratios and enhance throughput but also addresses broader issues like latency, jitter, and security vulnerabilities. This comprehensive solution is poised to revolutionize the management and performance of wireless networks, offering network administrators a powerful tool to ensure a seamless, efficient, and secure wireless communication environment. As the project progresses, the integration of machine learning models for predictive analysis will play a crucial role in anticipating potential issues and preemptively optimizing network performance. The user-friendly interface developed as part of the project will empower network administrators to visualize and interpret the analyzed data effectively, facilitating informed decision-making and swift responses to evolving network conditions.

II. LITERATURE REVIEW

[1]K-means clustering groups sensor nodes into clusters based on criteria like proximity or communication patterns, aiding in organizing them into manageable clusters. This facilitates the analysis and selection of suitable relay nodes from the pool of all sensor nodes in the network.The paper highlights hardware capacity as a crucial factor for relay node selection, indicating that not all sensor nodes qualify due to hardware limitations. K-means clustering incorporates this criterion, identifying potential relay nodes surpassing the specified threshold during the clustering process. K-means clustering optimizes relay node selection by considering criteria like hardware capacity, proximity to the base station, and network coverage. The algorithm iteratively refines cluster centroids, representing potential relay nodes, until convergence, ensuring effective fulfillment of specified criteria. It can handle a significant number of sensor nodes distributed across a wide area, enabling the identification of optimal relay nodes while considering hardware constraints and network performance requirements.

[2] Markov Models offer a versatile framework for modeling wireless channels affected by dynamic conditions, such as fading and interference, which are common in industrial environments. They provide a probabilistic representation of channel behavior over time, enabling the analysis of system performance and the design of optimal controllers. Markov models play a crucial role in wireless networked control systems (WNCSs) by capturing the stochastic nature of wireless channels. They help in understanding the dynamics of communication channels and enable the development of robust control strategies that can adapt to changing channel conditions. Markov models, several fundamental and important factors influence their effectiveness and applicability are State Space, Memory-lessness Property, Model Order, Training Data, Model Validation, Computational Resources, Transition possibilities ,Initial State Distribution.

[3] Time Series Analysis algorithms are vital in networking for understanding and predicting network behavior over time. Time series analysis algorithms analyze network parameters like traffic volume, bandwidth utilization, latency, packet loss, and performance metrics. They detect patterns, trends, and anomalies, aiding proactive network management and troubleshooting. Used in capacity planning, resource allocation, traffic prediction, and anomaly detection, they optimize network performance and facilitate efficient networking environments. The Time series analysis algorithms encompass statistical methods like ARIMA, exponential smoothing, and machine learning techniques such as neural networks and LSTM models. They accommodate various network data types, including time-stamped logs, SNMP data, flow records (e.g., NetFlow), and performance metrics from network devices and probes. These algorithms are applicable across different network stack levels, including the physical, network, and application layers, enabling signal analysis, traffic analysis, and user behavior analysis. Time series analysis algorithms are essential for network monitoring and management, providing insights into behavior and performance trends. They aid in congestion detection, traffic prediction, resource optimization, and ensuring QoS for critical applications. They identify anomalies and security threats, enhancing security and mitigating cyber-attacks. Time series analysis predicts future network trends, enabling proactive capacity planning and efficient infrastructure management. These algorithms support predictive maintenance, reducing downtime and improving overall network reliability.

[4]Random Forest algorithm is widely used in networking for various purposes due to its versatility and effectiveness. Random Forest algorithm is used in networking for intrusion detection, anomaly detection, traffic classification, performance prediction, and fault diagnosis. Random Forest algorithm enhances network security by accurately identifying and mitigating cyber threats, safeguarding network assets and data. Its ability to handle large-scale network data is crucial for analyzing complex environments and extracting actionable intelligence. Random Forest algorithm constructs an ensemble of decision trees, each trained on a random subset of data and features. Each tree predicts the target variable based on a subset of input features. Predictions from individual trees are combined through voting or averaging for final classification. Techniques like bootstrapping and feature randomness enhance generalization and reduce overfitting. In networking, Random Forest is trained on labeled datasets of network data samples and deployed for real-time analysis and decision-making.

[5] The Fourier Transform is a mathematical tool used to decompose complex signals into simpler sinusoidal components. It's widely utilized in various fields including wireless networking for signal processing, modulation, and demodulation purposes. The Fourier Transform allows for the analysis of signals in both the time and frequency domains, providing insights into the frequency components present in a signal. In wireless communication systems, different frequency components carry different types of information. The Fourier Transform enables the separation and analysis of these components, aiding in efficient signal processing.

It helps in understanding channel characteristics, identifying interference, and designing efficient modulation schemes to transmit data reliably. The Fourier Transform is vital in wireless networking:
Modulation and Demodulation: Enables conversion between digital data and analog waveforms, facilitating communication over analog channels.
Channel Estimation and Equalization: Analyzes channel frequency response, aiding in estimating characteristics and compensating for distortions via equalization.
Filtering: Designs filters using Fourier Transform to remove noise and interference, improving communication clarity and signal quality.
Frequency Allocation: Utilizes signal frequency characteristics for efficient allocation of frequency bands, minimizing interference and maximizing spectrum usage.

[6] The Hidden Markov model (HMM) is a probabilistic model that consists of two main components: hidden states and observed outputs and works on a transition matrix. 1. Hidden States: These are unobservable variables that form a Markov chain, meaning the probability of transitioning to the next state depends only on the current state. 2. Observed Outputs: Each hidden state emits observable outputs with certain probabilities, known as emission probabilities.
The working of an HMM involves three fundamental tasks:
1.State Transition: Given an initial distribution over the hidden states and transition probabilities between states, the model moves from one hidden state to another according to the transition probabilities.
2.Observation Emission: At each hidden state, the model generates an observable output based on the emission probabilities associated with that state.
3.Inference: Given a sequence of observed outputs, the task is to infer the most likely sequence of hidden states that could have generated those observations. Training an HMM involves estimating the model parameters (transition probabilities and emission probabilities) from a set of training data. HMMs are used in various applications such as speech recognition, natural language processing, bioinformatics, and time series analysis due to their ability to model sequential data with hidden information.

[7] The Random Walk model is a mathematical concept used to describe the movement of an entity that takes random steps in a defined space and works on a transition matrix. Here's how it works:
1.Initialization: The random walk starts at a specific position in a space, which could be discrete or continuous
2.Random Steps: At each time step, the entity takes a step in a randomly chosen direction. The step size and direction are determined by a probability distribution. Common distributions include uniform, Gaussian, or other symmetric distributions.
3.Accumulated Movement: The entity continues taking random steps for a specified number of time steps or until a stopping condition is met. As it moves, it accumulates its position over time.
4.Termination: The random walk may terminate after a certain number of steps, upon reaching a boundary, or based on other predefined conditions.
5.Analysis: After the random walk is completed, statistical analysis may be performed on the accumulated positions. This analysis could include calculating properties such as mean displacement, variance, probability of returning to the starting point, or other relevant metrics.
The behavior of a random walk depends on the specific rules governing the random steps and the characteristics of the underlying space. Random walks have diverse applications in various fields, including finance, physics, biology, and computer science. They are used to model phenomena such as diffusion, stock price movements, molecular motion, search algorithms, and more.

II. SYSTEM OVERVIEW

3.1 Flow Chart

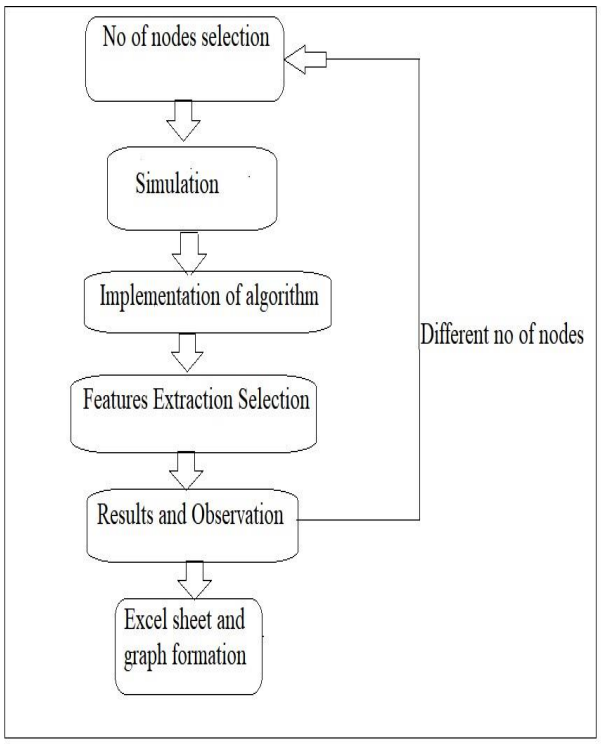

Figure. 1 Flow Diagram

The diagram shows the steps of a simulation process. The goal of the simulation is to extract features and observe the results for different numbers of nodes

III.METHODOLOGY

1. Number of Nodes Selection: Determine the size of the network by selecting the appropriate number of nodes to be monitored. This step is crucial for capturing a representative sample of network traffic and ensuring comprehensive analysis.
Number of nodes selected are 20 nodes, 50 nodes, 100 nodes, 250 nodes, 500 nodes.
2. Implementation of Algorithm: Choose and implement a suitable traffic analysis algorithm, such as packet sniffing or flow-based analysis. This step involves deploying the algorithm within the network infrastructure to capture and process data effectively.
3. Simulation: Simulate network traffic using the implemented algorithm to mimic real-world scenarios. This step helps evaluate the algorithm's performance under various conditions and ensures its effectiveness in different network environments.
4. Feature Extraction Selection: Identify and select relevant features to extract from the captured traffic data. These features could include packet size, source and destination addresses, protocol types, etc. Choosing appropriate features is essential for gaining insights into specific aspects of network behavior.
5. Result Observation: Analyze the extracted features to observe patterns, anomalies, or trends in the network traffic. This step involves interpreting the results of the traffic analysis algorithm to draw meaningful conclusions about the network's performance and potential issues.
6.Excel Sheet and Graph Formation: Organize the observed results in an Excel sheet for systematic documentation. Create graphical representations, such as charts or graphs, to visually illustrate trends and variations in network traffic, making it easier to communicate findings and understand the data at a glance.
7. Different Number of Nodes: Repeat the entire process with varying numbers of nodes to assess the scalability and performance of the network traffic analyzer. This step helps understand how the system handles different network sizes and ensures that the analysis remains effective as the network scales.

The same procedure is carried out for 7 different algorithms 1.K means Clustering 2.Random Forests 3.Markov Model 4.Fourier Transform 5.Time Series Analysis 6.Hidden Markov Model 7.Random Walk

As Markov Model, Hidden Markov Model, Random Walk works on transition matrix all 3 algorithms are applied on the same topology for different numbers of nodes to analyze the efficiency of each algorithm. Performance graph is plotted on the output values.

IV. RESULTS AND DISCUSSION

	Algorithm	Drop Ratio	Throughput
20 nodes	K means clustering	0.16216	0.83783
	Random Forest	0.5	0.5
	Markov model	0.07999	0.92
	Fourier Transform	0.205128	0.794871
	Time Series Analysis	0.199999	0.8
50 nodes	K means clustering	0.17073	0.829268
	Random Forest	0.2888	0.711111
	Markov model	0.110638	0.88936
	Fourier Transform	0.19385	0.80614
	Time Series Analysis	0.14	0.86
100 nodes	K means clustering	0.21568	0.78431
	Random Forest	0.20895	0.79104
	Markov model	0.09989	0.9001
	Fourier Transform	0.20607	0.79392
	Time Series Analysis	0.14	0.86
250 nodes	K means clustering	0.207336	0.79266
	Random Forest	0.20389	0.7961
	Markov model	0.09798	0.90201
	Fourier Transform	0.20038	0.799614
	Time Series Analysis	0.191999	0.808
500 nodes	K means clustering	0.19947	0.80052
	Random Forest	0.19959	0.8004
	Markov model	0.09923	0.90076
	Fourier Transform	0.198329	0.80167
	Time Series Analysis	0.201999	0.798

Figure 5.1 Outputs of each algorithm in tabular format.

Figure 5.1 Shows drop ratio and throughput given by each algorithm for a different number of nodes.

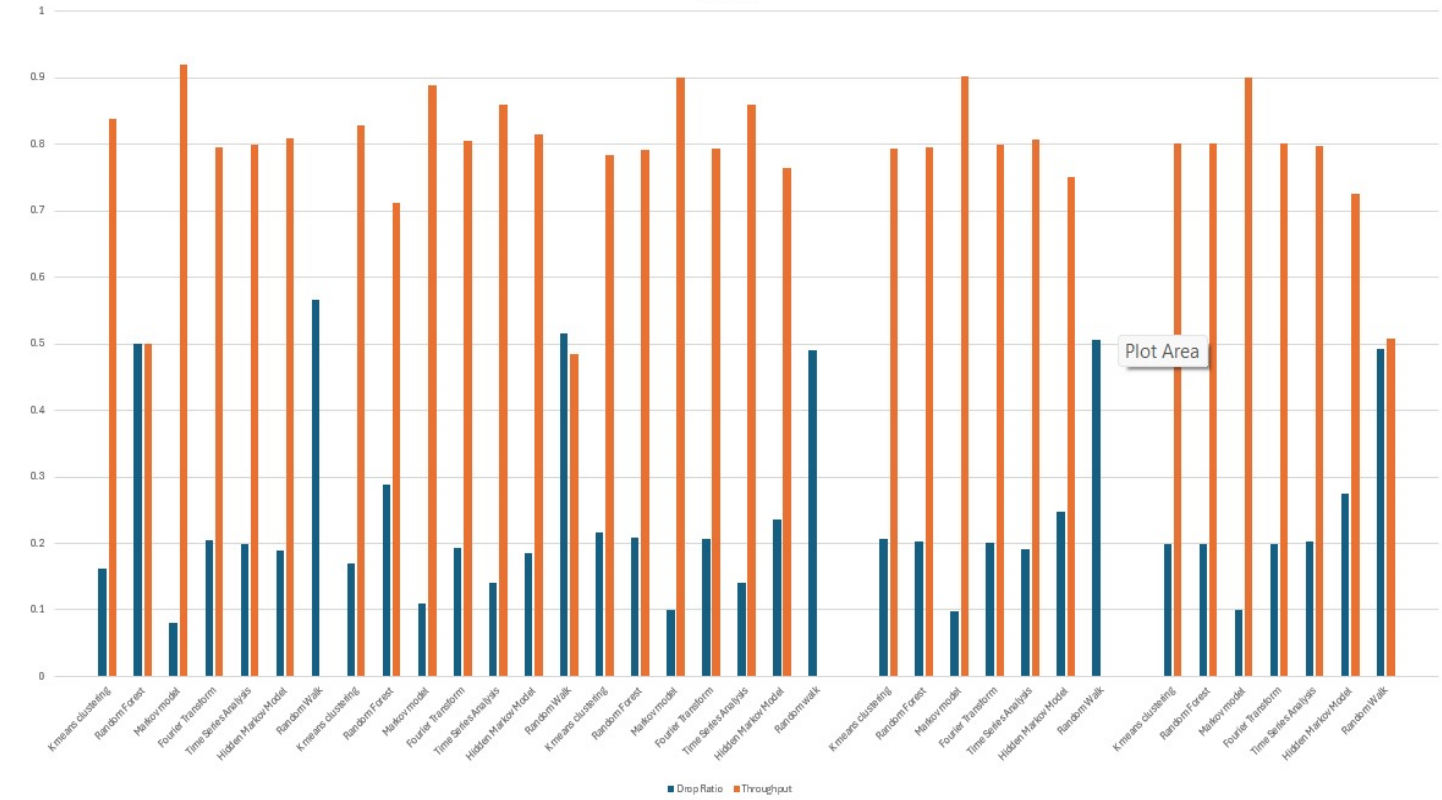

Figure 5.2 Chart for performance analysis

Figure 5.2 Output of each algorithm is represented in the form of a chart which helps in comparing performance of each algorithm for different numbers of nodes.

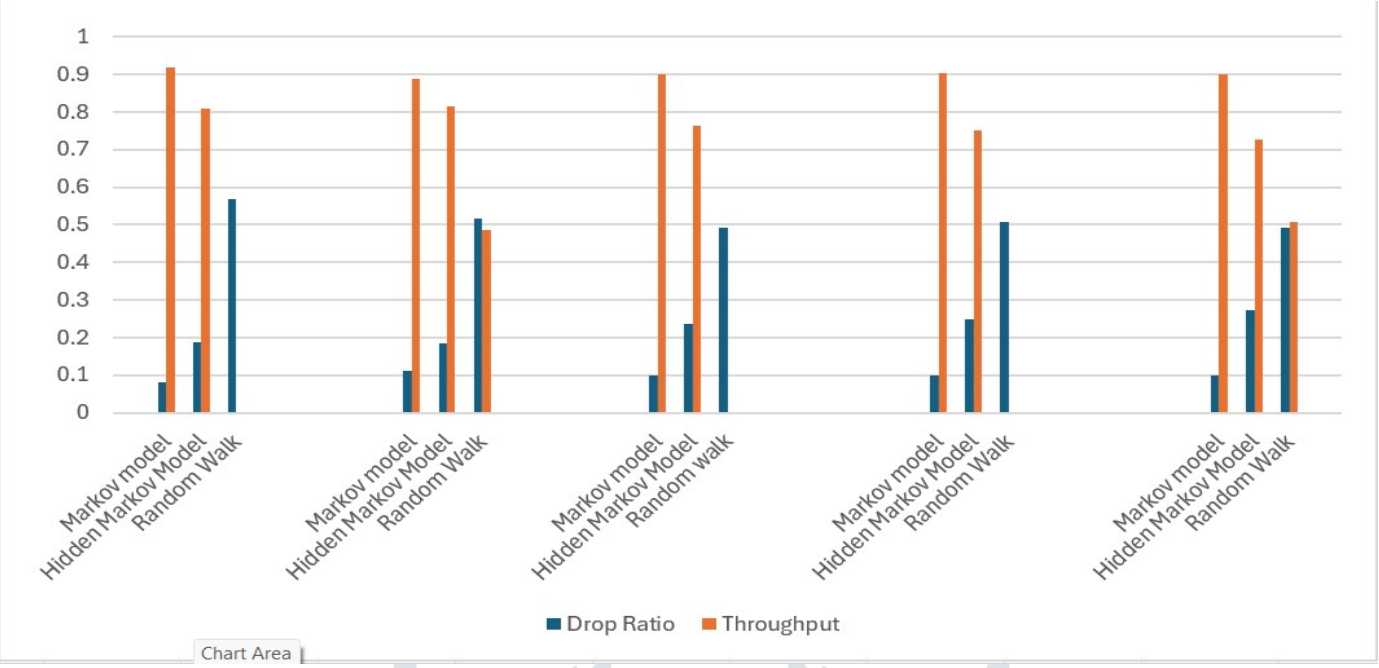

Figure 5.3 Chart for performance analysis of algorithms working on transition matrix.

Figure 5.3 Output of 3 algorithms working on the transition matrix is represented in the form of a chart which helps in comparing performance of each algorithm for different numbers of nodes.

VI. OUTCOMES

1.Gaining insight into the relevant algorithm used in traffic analysis.

2.Checking the performance of the algorithm for different input and conditions.

3.Resulting in the optimization factor for different parameters of a network.

VII. CONCLUSION

Average drop ratio of average drop ratio is 9.55 percent and average throughput is 90.055 percent is represented in Figure 5.1 and Figure 5.2.

From figure 5.2 and figure 5.3 performance of Markov Model algorithm is efficient compared to K means Clustering, Random Forests, Markov Model, Fourier Transform, Time Series Analysis, Hidden Markov Model, Random Walk.

Markov Model shows considerable throughput and drop ratio compared to other algorithms in the same simulator environment.

references

[1] Zhang, Y., et al. (2017). "Network Traffic Classification Using K-Means Clustering." In IEEE Transactions on Information Forensics and Security

[2] Wu, Y., et al. (2015). "A Markov Chain Model for the Bursty Characteristics of Network Traffic." In IEEE Transactions on Network and Service Management

[3] Zhang, X., et al. (2019). "Time Series Analysis of Network Traffic for Anomaly Detection." In Proceedings of the ACM SIGCOMM Conference K-Means Clustering in WSN uploaded by Sergio Andres Diaz Salas.

[4] Lee, Y., et al. (2016). "Traffic Classification and Identification Using Fast Fourier Transform." In Journal of Network and Computer Applications.

[5] Zaman, R. U., et al. (2018). "Random Forests-Based Classification of Network Traffic.

[6].Data Aggregation & Transfer in Data Centric Network Using Spin Protocol in WSN by Prathima Kadlikoppa[1], Indira Umarji[2], Sachin Patil[3] [1]Student, SDM College of Engineering & Technology, Dharwad [2]Asst. Professor, SDM College of Engineering & Technology, Dharwad [3]Asst. Professor, SKSVMACET, Lakshmeshwar

[7] HIDDEN MARKOV MODEL TRAFFIC CHARACTERISATION IN WIRELESS NETWORKS Evangelos D. Spyrou and Dimitrios K. Mitrakos School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Egnatia Odos, Panepistimioupoli, Thessaloniki, Greece

[8] Random Walks with Jumps in Wireless Sensor Networks Leonidas Tzevelekas University of Athens Greece Email: ltzev@di.uoa.gr Ioannis Stavrakakis University of Athens Greece Email: ioannis@di.uoa.gr