Sustainable Next Generation Network Design using Social Aware and Delay Tolerant Approaches

Ambreen Memon

A Thesis Submitted to
Auckland University of Technology
in Fulfilment for the Degree of
Doctor of Philosophy

May 2021

School of Engineering, Computer, and Mathematical Sciences

Abstract

The Internet of Things (IoT) perspective endeavours to unite all the physical objects or ‘things’ embedded with electronics, software, sensors, and network connectivity to allow more direct integrations between the physical world and cyber-based systems. At the same time, these networked devices and associated data communications can increase the energy demands exponentially and could seriously harm the environment because of their carbon footprints. Currently, most researchers and practitioners have dedicated their efforts to improving the energy efficiency of IoT systems. But most are falling into a box of focus within telecommunications and Internet domains themselves. This research targets to develop a novel, eco-friendly and sustainable data transmission approach to offload the data traffic from infrastructures to opportunistic and social networking by exploring the existing movements and spatial closeness relations among things. Objectively, to complement the traditional infrastructure-based data transmission, the new idea is to optimally piggyback the data on the moving physical objects/things for the data delivery to achieve energy reduction and ensure their Quality of Services (QoS) requirements.

The thesis contribution is presented in three phases; in the first phase, the idea of similarity in the user’s mobility patterns is explored. These similarity patterns are found using the co-occurrence matrix. The identified patterns further motivate the second phase, where the research predicts encounters among mobile users. This enables the approach to investigate different mobility patterns and find common points of meeting among other users. The meeting points are then predicted using the random forest approach. After getting the accurate results of ‘prediction of meeting (encounter)’. The multiple-criteria decision model will decide based on delay and transfer the data through the Internet or Device-to-Device (D2D) in the third phase.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

API	Application Programming Interface
D2D	Device-to-Device (D2D)
DTN	Delay Tolerant Network
EEDDA	Energy-efficient data dissemination Approach
ICTs	Information and Communication Technologies
IoT	Internet of Moving Things
IPs	Internet Provider Security
LAN	Local Area Network
MAC	Media Access Control Address
MAN	Metropolitan Area Network
MANET	Mobile Ad-hoc Network
MANET	Mobile Ad-hoc Network
QoS	Quality of Services
SAA	Similarity Analysis Approach

Attestation of Authorship

I hereby declare that this submission is my work. To the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Ambreen Memon

May 2021

List of Publications

Ambreen Memon, William Liu, and Adnan Al-Anbuky' A New Energy Efficient Big Data Dissemination Approach Using the Opportunistic D2D Communications' Third International Conference, SmartGIFT 2018 Auckland, New Zealand, April 23–24, 2018.

Ambreen Memon, William Liu, and Adnan Al-Anbuky 'CatchMe If You Can: Enable Sustainable Communications Using Internet of Movable Things' 2016 IEEE 14th Intl Conference on Dependable, Autonomic and Secure Computing, 14th Intl Conference on Pervasive Intelligence and Computing, 2nd Intl Conference on Big Data Intelligence and Computing and Cyber Science and Technology.

Manuscripts for future publications :

Ambreen Memon, Jeff Kilby and M. Ishtiaq 'Sustainable Smart Connected Buildings: Random Forest-based Energy Saving Model Elsevier Securing IoT-based Critical Infrastructure journal 2021.

Ambreen Memon, Jeff Kilby, and M.Ishtiaq'Analysis and Implementation of Human Mobility Behaviour using Similarity Analysis Based on Co-occurrence Matrix'. Journal Paper.

Ambreen Memon and Jeff Kilby 'Energy-Efficient Data discrimination Approach.' Journal Paper.

Ambreen Memon and Jeff Kilby 'An Efficient Approach For Big Data Transmission Using The Encounter Prediction.' Journal Paper.

Acknowledgement

I would like to thank both my supervisor's Dr Jeff Kilby and Professor Jairo Gutierrez, for their steadfast integrity and selfless dedication. My primary supervisor, Dr Jeff Kilby, for his support, enthusiastic encouragement, and guidance to accomplish the dissertation. He is an excellent supervisor; he supported and encouraged me to develop the skills and improve my performance. My secondary supervisor, Professor Jairo Gutierrez, provided informative discussions, suggestions, manuscript reviews, and valuable critiques on my research work.

I would like to convey my deepest gratitude to all the Network and Security Research Group (NSRG) at Auckland University of Technology. It has been a source of networking (friendship), good advice, and collaboration.

I would like to express my deepest gratitude towards my supportive husband, Abdul Qadir, who always motivating and encouraging me; without his support, I wouldn't have achieved this milestone.

I would also like to thank my family, uncle like my father Abdul Jabbar Memon, and children Aamna, Aaliya, and Muhammad Ali; my brother, sister, and cousin supported me on this journey.

Last, I wish to thank my best friend Rashmi Munjal for her support and motivation.

Chaper 1
Introduction

1.1 Background

The past decade has witnessed a significant proliferation of Internet-capable devices [1]. While its most important commercial impact has been in consumer electronics, with the smartphone revolution and the uptake of wearables, connecting humans is only part of a more significant trend toward the interconnection of the physical world with the digital world [1]. The expansion of Information and Communications Technology (ICT) has provided rich data sources for analysing, modelling, and interpreting human mobility patterns. Many researchers already introduced behaviour-aware protocols to understand better architecture and realistically model behavioural characteristics [2], similarity [3], and aggregation of mobile users. These models are essential for potential analysis professional development and computation. Jahromi et al. [4] show the utilisation of mobile phone data for transport models, while the Internet is a communication network connecting people to information.

The Internet of Things (IoT) is a network that connects different devices (or things), where things are wirelessly connected via smart sensors [5-7], shown in Figure 1.1.

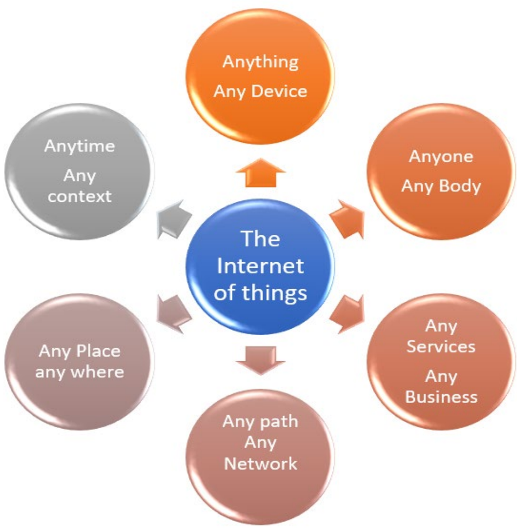


Figure 1.1 Internet of Things

The words *'Internet'* and *'Things'* mean an interconnected worldwide network based on sensory, communication, networking, and information processing technologies, which might be the new version of Information and Communications Technology (ICT) [8]. IoT has been defined as a dynamic global network infrastructure with self-configuration capabilities based on standards and interoperable communication protocols. Physical and virtual *'things'* in an IoT have identities and attributes. They are capable of using intelligent interfaces and being integrated as an information network [9]. The *'things'* in IoT networks can be read, recognised, located, addressed through information sensing devices, or controlled via the Internet, irrespective of communication, radio frequency identification (RFID), wireless local area networks (WLANs), wide area networks (WANs) [7]. An IoT is integrated with billions of objects that can sense, communicate and share information. All devices are interconnected over public or private Internet Protocol (IP) networks. These interconnected objects have data regularly collected, analysed, and used to initiate action, providing a wealth of intelligence for planning, management, and decision making; this is the world of the IoT [10]. The Internet is not only a network of computers, but it has evolved into a network of the device of all type and sizes, vehicles, smartphones, home appliances, toys, cameras, medical instruments and industrial systems, animals, people. For example, buildings can communicate and share information based on stipulated protocols to achieve smart reorganisations, positioning, tracing safe control, and even personal real-time online monitoring, online upgrade, process control and administration [7]. IoT utilises the aggregated impact of these organised objects for improved efficiency, accuracy, and economic benefits for humans [11]. A smartphone may opportunistically pass data to another smartphone in a forward manner or as a direct destination or use it as a carrier to give the data to a third smartphone. This network of connectivity may be referred to as an opportunistic network. Routing algorithms built for these networks aim to increase the probability of effective transmission of messages. The most common method is to measure the likelihood of transmitting a message using details such as node contacts and information of the area while using past experiences to forecast potential ones. Conventional mechanisms for routing are not suitable [12].

1.1.1 Characteristics of Internet of Things

The fundamental characteristics of the IoT are as follows:

Interconnectivity – about the IoT, anything can be interconnected with the global information and communication infrastructure [7].

Things-related Services – the IoT can provide thing-related services within the constraints of things, such as privacy protection and semantic consistency between physical things and their associated virtual things. So, to deliver thing-related services within the limitations of things, both the physical and the information world technologies will change [7].

Heterogeneity – the devices in the IoT are heterogeneous based on different hardware platforms and networks. They can interact with other devices or service platforms through various networks [7].

Dynamic Changes – the state of devices changes dynamically, for example, sleeping and waking up, connected or disconnected, and the context of devices, including location and speed. Moreover, the number of devices can change dynamically [7].

Enormous Scale – the number of devices that need to be managed and communicate with each other will be at least an order of magnitude larger than the devices connected to the current internet [7].

Based on the network topology, it can comprehensively classify system association into two types:

1. Infrastructure empowered association and specially appointed.
2. Astute association of infrastructure-less.

The first type uses the existing infrastructure, e.g. mobile base stations, access points, switches. At the same time, the second utilises the infrastructure-free, short-run radio systems, e.g. Bluetooth, Wi-Fi, and so forth, to fabricate decentralised and commercial ad-hoc systems [11]. The device connected to IoT can share data using different methods, and Mobile Crowdsourcing is one of them [13].

1.1.2 Mobile Crowd Sensing

Mobile Crowd Sensing (MCS) represents an IoT application category that relies on data collected from mobile sensing devices such as smartphones [14]. Compared with the traditional mote-class wireless sensor networks, MCS's unique advantages include:

1. Most mobile devices have significantly more storage, computation, and communication resources than mote-class sensors. They have several communication interfaces, e.g. Wi-Fi, Near-Field Communication (NFC), and Bluetooth, inherently equipped with multimodality sensing capabilities.
2. Mobile devices avoid the cost and time of deploying large-scale wireless sensor networks because people carry mobile devices wherever they go and whatever they do. These features make them ideal candidates to carry out the delicate data transmission task in the Internet of things domain.

1.1.3 Problems with Existing Ad-hoc Mobile Networks

High energy consumption mobile ad-hoc networks use radio Base Stations (BSs) or Access Points (APs) to Mobile Terminals (MTs) with the core network or the Internet [15]. The BS is the primary source of energy consumption in the wireless access network from the operator side. It is estimated that more than 57% of the operator's total energy consumption is in the BS [11]. There are about three million BSs worldwide that consume 4.5 GW of power [16]. It is estimated that three billion MTs globally have a power consumption of 0.2 to 0.4 GW [17]. The high energy consumption of wireless access networks has promoted increased environmental and financial concerns for both service operators and users and Quality-of-Experience (QoE) considerations for mobile users. When BSs are suffering from *'over-provisioning problems'* [17], by tuning the BSs, on or off, based on traffic load dynamics, energy can be significantly saved from the perspective of mobile networks or Mobile Ad-Hoc Network (MANET).

A mobile ad-hoc network collect nodes that act independently but are indirectly dependent on each other for their operation in the network. These networks work on a mobile node's limited battery capacity, affecting network survivability since links are disconnected when the battery is exhausted [18]. In case of disconnection from the network, the node will not sense and process the data [19].

1.1.4 Routing

Because mobile ad-hoc networks are defined by a multi-hop network topology that can frequently alter due to volatility, efficient routing protocols must create communication paths between nodes without creating undue overhead traffic management or computing pressure on the power-restricted devices. Simulation studies have shown that routing protocol efficiency in latency, packet failure, delay, and overhead regulation depends heavily on network conditions such as traffic load, agility, density, and the number of nodes. Existing directory-based infrastructure and resource-discovery systems such as Universal Plug and Play (UPnP) cannot manage the complexities of ad-hoc networks [20]. There is currently no established approach, but it is evident that these protocols should be built in close cooperation with the routing protocols and should include context-awareness (location, neighbourhood, user profile, etc.) to improve performance. Although ad-hoc networks are often linked to fixed infrastructure (e.g. telephone, wireless network, etc.), protocols and methods are required to enter the ad-hoc network of accessible external networks provided by service providers and content providers.

1.1.5 Addressing and Internet Connectivity

So, to allow nodes to interact inside the ad-hoc network, every node requires an address. The usage of IP addresses is unnecessary for stand-alone ad-hoc networks, as separate Media Access Control (MAC) addresses may be used to access nodes. However, due to node versatility, overhead factors, and other options for address distribution, an internal address structure of prefixes and ranges as in the fixed Internet is hard to manage in mobile ad-hoc networks [21].

The studies clarify that although several approaches are being investigated, there is still no widely accepted approach for addressing Internet access. Current examples of host methods identity where the IP's function is restricted for routing and not addressing [20].

1.2 Research Gap

Green networking [11, 22] refers to the information to update networking or make the communication more powerful with any procedure that finally decreases energy consumption. Because of the possible economic benefits and the expected environmental benefits, reducing excessive energy usage is becoming a significant concern in wired/wireless networking. The impact of these concerns is often referred to as green networking.

In a Delay Tolerant Network (DTN), this research defines the source reducing the Network lifetime. Recognising the importance of reducing energy consumption, advanced this research and investigate the sources that consume energy. Green networking has various basic applications, especially as energy becomes more of a concern and people become more conscious of the adverse effects of energy use. Thus green networking attracts the attention of analysts. A DTN, this research delineates the sources, reducing the network lifetime.

Understanding the significance of the need for reduced energy consumption, the investigation of the sources has advance that expends energy unfruitful. A DTN [23, 24] is portrayed as a network with regular discontinuous associations, disturbances, and experience since a long time ago deferred information conveyance due to nodal mobility, inadequately sent nodes, node disappointments, etc. Be that as it may, the nodal development has been used for the information communication in such networks, which is avoided as Store-Carry and Forward [25, 26].

Before investigating the energy consumption factors, let's first solidify the applications and the necessity for energy administration plans. This literature helps understand the demand for new technology, which must be effective in all three concepts. This research introduces a new idea to reduce the energy and improve the Delay Tolerant Network with Social Layer (Similarity Matrix) while considering all the above factors.

This research carried and presented in this thesis aims to answer the following questions:

Question 1: What can be new advancements for energy efficiency in the opportunistic network?

The proposed research focuses on energy efficiency while sustaining Quality of Service (QoS) in the heterogeneous network (Infrastructure/Infrastructure-less Network). The literature survey concludes that there are many ways to use traditional communication with massive energy consumption and network bandwidth congestion. Our work will focus on utilising resources to reduce traffic volume and reduce energy consumption using Device-to-Device communication. Suppose the distance between two nodes is diminished with the concept of D2D. In that case, the QoS can be sustained while determining the sustainability challenges such as delivery probability and latency in heterogeneous networks. Distance is directly proportional to energy consumption. If the distance is increased, energy will gradually increase. In this work, devices will communicate according to their social relationship and contact each other at a location, exchanging the data with their delay-tolerant features and encounter-time. But, if they do not meet specific requirements, they will use the traditional communication method, including an infrastructure-less network, which aims to sustain QoS.

Question 2: How can user mobility similarity matrixes help implement the idea of D2D to know the exact mobility patterns to exchange data?

To initiate D2D, it is essential to know when two users will be at the same location and time. It is observed the user's similarity pattern as per their routine movements and schedule defined. There are many factors in real life, which affect every user's daily movement patterns. The model must consider their MAC address, IP address, location, and time to be in the same range for communication. Each node analysed the time and location of the next node, when and where it will meet him next, and forward the same message accordingly. These decisions are based on real-life traces of nodes and their social relationship of nodes according to their daily routines, workplaces, etc.

Question 3: ***How will the encounter patterns extracted from the similarity matrix predict future movement to enable data communication?***

The similarity matrix resulted in the user's encounter patterns. However, the model must analyse their future movement patterns from the previous history to enable data communication. It is essential to examine the history of encounter patterns and apply some algorithms to predict future movements to address this issue. This will help us know that users will contact each other for how long and what time. In such a way, the connection will be established to exchange data; otherwise, if there is less or no probability to meet within delay hours, data will be uploaded onto the Internet.

Question 4: ***How much energy can be saved while transmitting data using D2D communication compared to traditional networks?***

The research is focused on designing an energy-efficient data dissemination architecture, but how much energy can be saved while using the D2D. To address this issue, it is essential to compare the energy consumption in traditional networks and D2D to utilise them efficiently for data dissemination in case of multi-decision for data volume, delay, and encounter time.

The following section lists and discusses the contributions of the thesis.

1.2.1 Contributions

This thesis develops an eco-friendly and data transmission approach that explores the existing movements and spatial closeness relations among *'things'*.

Contribution 1: This research has established a relationship between distance and energy consumption of nodes. It is identified that the data transmission distance is a key but changeable factor contributing to the overall energy consumption. It is straightforward that the longer the transmission distance is, the higher the total energy consumption. In our first contribution to design, energy-efficient architecture, the similarity of user patterns will be analysed at the 1st stage. A Similarity-Analysis Approach (SAA) model is designed to determine user mobility behaviour as per their spatial-temporal preferences such as MAC address, IP address, physical location, and time. The user's similarity pattern can be observed as per their routine movements and schedule defined. A similarity matrix is outlined for all users of each location and time from the USC dataset to validate the model. The SAA approach will generate a co-occurrence matrix based on users' similar behaviour and calculate their encounters for a location/time. The co-occurrence matrix is used for capturing user's mobility behaviour and preferences. A mobile encounter index metric is also proposed to indicate the similarity ranking between two objects' mobility preferences. A technique is also developed for producing a co-occurrences matrix of users based on their similar behaviour to determine their encounters. Our results show that our proposed approach SAA is 33 % more efficient than the previous similarity approach.

Contribution 2: After analysing user's encounter patterns, it is essential to know their future prediction to exchange data. Therefore, a random forest model is proposed, which used the mobility traces of different users for future location prediction and then employed these future locations for encounter prediction. The random forest model is based on the decision trees that store the mobility history of each user. Whenever a user visits a new location or an already visited location at a different time cycle, the tree is updated to reflect the new pattern. Moreover, hyperparameter tuning is used in the random forest to enhance the predictive strength and fasten the model. The USC mobility traces are used in the dataset for the experiments. The purpose of our proposed model is to reduce resource consumption, i.e. energy and bandwidth during data transmission in daily movements. The Gaussian process and random forest are compared and analysed using real-world mobility data traces to predict an accurate encounter pattern.

Contribution 3: The research proposes an energy-saving model for multiple-criteria decision device-device communication using Device to Device (D2D) and the Internet, which helps sustain the QoS. An Energy-Efficient Data Dissemination Approach (EEDA) is designed, which will take multiple-criteria decisions by using the encounter probability results concerning the delay time of the user, the transfer time of data, encounter time of senders and receivers, the data size. The model presented will decide that the data may be transferred through the Internet or D2D. So, to reduce the energy consumption, the current system first uses the USC data traces of a user having a mobile phone (node) with an encounter device roaming at the location. Energy consumption will be calculated in both networks. After that encounter, the device forwards them to the decision engine. With the help of the decision engine, data will go through the Internet and sustain the QoS or Data travel D2D and save energy consumption.

1.3 Methodology

The methodology of this thesis consists of modelling and simulations. At first, an energy consumption model was computed, followed by the proposal of a co-occurrence matrix for capturing human mobility and preference patterns. A given object's co-occurrence matrix M can be constructed to identify the primary mobility trend or mobility preferences. Furthermore, a mobile encounter co-occurrence metric is also proposed to indicate the similarity ranking between two objectives' mobility preferences by quantitatively measuring the similarity between their mobility matrices' eigenvectors. At first, it was validated if the mobile encounter phenomena exist in mobile users' society by obtaining the similarity distribution among mobile peers. The simulation studies showed the correlations between the overall network performance and the mobility similarity among users. Simulations were performed on the datasets and mobility information which are available from the USC mobility traces repository. The datasets were first Analysed and pre-processed. These simulations showed the correlations between the mobility similarity threshold, the next hop of the node to route and forward data, and the overall network performance, such as success data deliver ratio. A similarity ranking distribution histogram is also prepared for the datasets used to show the number of users as a function of similarity rank that quantifies the mobility similarity between peers. For reducing energy consumption, the system first uses the mobility traces of a user having a mobile phone (node). These mobility patterns yield the time-space associations of the sender and receiver nodes. The time-space associations of both nodes are used for similarity ranking. Based on similarity profiles, the future location and encounter time interval are predicted for different users. The energy consumption model based on the opportunistic model works to reduce energy consumption in the overall network in the last phase. The random forest model for encounter prediction consists of a set of random regression trees. Decision trees possess the decision-making ability based on the results of a specific condition. The Random forest model is evaluated for encounter prediction among users. The evaluation is done step by step. At first, the results for the encounter were calculated using both Gaussian and Random forest approaches. After this, the encounter prediction results are computed with both approaches, and a comparison among the results obtained through both approaches was performed. Finally, based on the Random Forest Model results, the Energy Efficient Data Dissemination Approach (EEDDA) is used to calculate the energy comsumption.

1.4 Thesis Structure

This overall thesis structure, shown in Figure 1.2, starts with an introduction to relevant background knowledge of IoT devices, contributions of research, and research questions in the first chapter. Chapter 2 presents a detailed Literature Review.

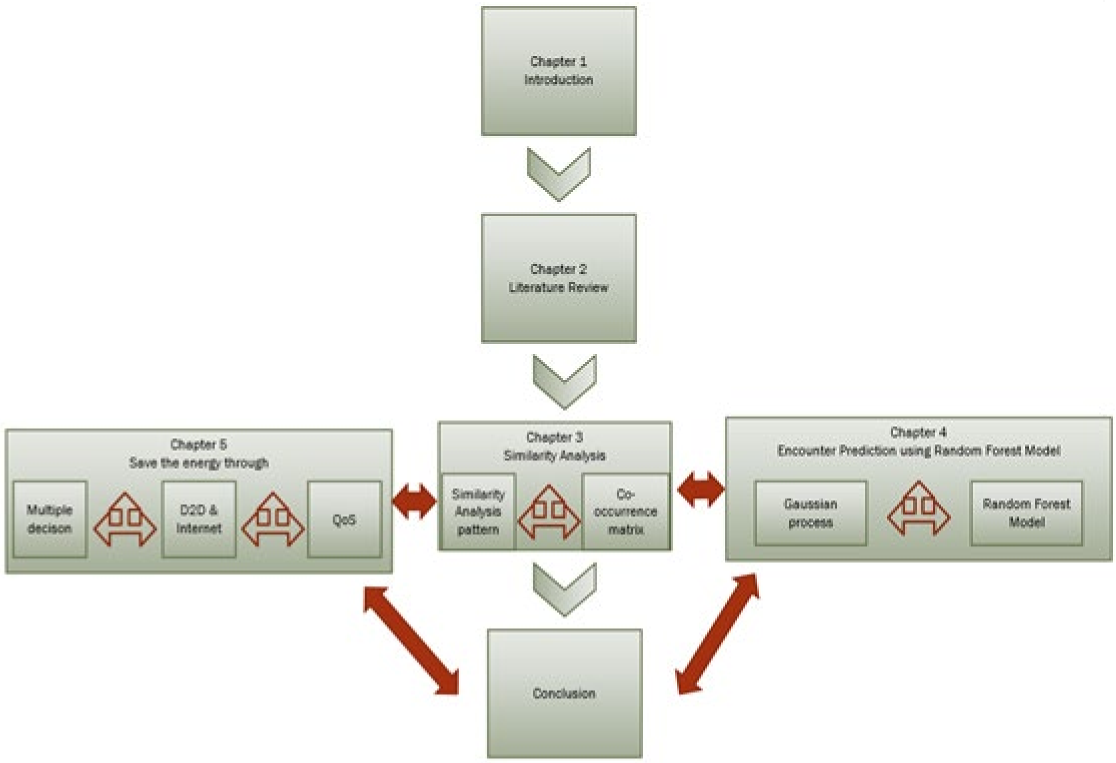


Figure 1.2 Thesis Structure

The structure of chapter 3, shown in Figure 1.3, presents a similarity analysis in which a relationship is established among nodes distance and energy consumption.

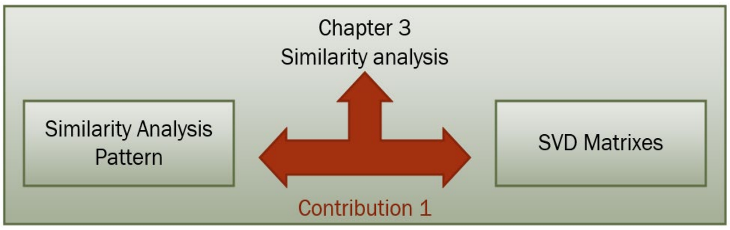


Figure 1.3 Chapter 3 structure

The data transmission distance is a key but changeable factor contributing to the overall energy consumption. It is straightforward that the longer the transmission distance is, the higher the total energy consumption. Firstly, it consists of a background that gives a complete overview of Characteristics of Human Mobility Behaviour, Skewed location visiting preferences, Mobility Heterogeneity, Mobility Correlations, IP address, and MAC address. This chapter gives a brief overview of the similarity analysis pattern. The remaining chapter follows the system model for similarity, co-occurrence matrix, similarity analysis, Case Study, Dataset, Location information, Data Gathering, Numerical Studies, and Conclusion. In this chapter, the proposed approach utilises device info, i.e. IP and MAC address. The proposed scenario, where various individuals are in the same building but are marginally distant from each other, i.e. separate floors. So the proposed solution assumes all users have the same accessibility nature rather than having specific data communication networks. The same network helps to show specific mobile behaviours. The SAA is provided for finding the similar behaviour of users' mobility. It is the most suitable and cost-efficient way of analysing mobility behaviour.

The structure of Chapter 4, shown in Figure 1.4, shows the detail of encounter prediction using the random forest model.

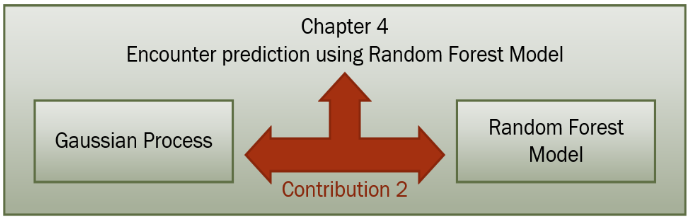


Figure 1.4 Chapter 4 structure

In this chapter, the Gaussian process and random forest are compared and analysed using real-world mobility data traces and show how these encounter prediction results save more energy. This chapter also proposes using supervised learning techniques together with Random Forest Regressor and Gaussian process modelling. To predict future encounters based on historical patterns of individual nodes to compare the findings produced from the Gaussian and Random Forest model, the precision of predicted encounters at various locations is measured by all methods. Such tests contrasted the accuracy of predicted encounters in separate weeks of the month at a similar position. For our experiments, the USC data traces are used. The traces are actual measurements taken from the university campus of USC. The data contains information about Wi-Fi associations, user profiles. The data is about six different buildings on the campus, access points, data about all days, including weekdays and weekends.

The structure of Chapter 5, shown in Figure 1.5, presents an energy-efficient data discrimination approach using multiple-criteria decisions.

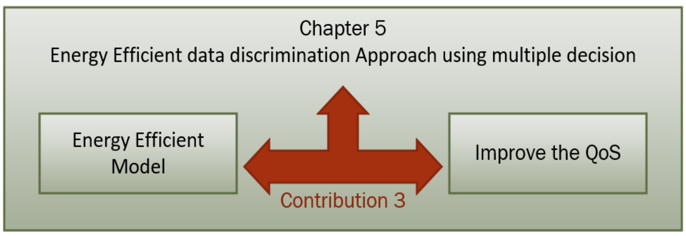


Figure 1.5 Chapter 5 structure

In this chapter, firstly, the System model for Energy Efficient approach is presented. Then data transmission through the internet network is explained. A comparison analysis is performed for energy consumption for device-internet and device-device. The University of Southern California (USC) mobility traces are used in the dataset for the experiments. Our proposed model aims to reduce resource consumption, i.e. energy and bandwidth during data transmission in daily movements achieved in this chapter. Hence our proposed approach improves the QoS as well. Chapter 5 concludes all the contributions and the findings in this thesis.

Finally, chapter 6 discusses the findings from the presented research in this thesis and future work.

Chaper 2
Literature Review

2.1 Introduction

The number of mobile users is rising rapidly in the emerging innovative digital world. Mobile devices form a Mobile Ad-Hoc Mobile Network (MANET) that offers different scalability and flexibility advantages. Other users can enter or leave the network without prior notice [27]. The mobility of mobile nodes allows them to communicate dynamically and arbitrarily, and user agility makes the MANET environment unpredictable for real-time data transfer applications [28].

Future social networks are expected to have groups of applications (apps) aware of mobile users' behavioural profiles and interests and likely embrace mobile peer-to-peer networking, including Delay Tolerant Networks (DTNs) [29]. In addition to behaviour-aware communication paradigms, such as participatory sensing, a new generation of protocols is evolving [30]. This behaviour-aware communication model controls user behaviour and expectations to achieve successful DTN activity, e.g., forwarding interest-based target messages, encounter-based routing, exploration of mobile resources. For the research, performance assessment, and simulation of such networking protocols, accurate models of mobile user behavioural profiles are essential. Therefore, there is a compelling need to recognize and model behavioural profiles, similarities, and user clustering of mobile users in a practical way [31, 32].

Lindgren, Schulen, and Doria [33] introduced the Prophet-Probabilistic Routing Protocol History of Encounters and Transitivity (PRoPHET), which has been implemented using encounter history and transitivity. This protocol's structure to transmit a message is structured to exploit transitivity and historical encounter. It implies that the nodes have a repeated mode that is predictable but not a random movement. Such a move will result in the use of a fast and efficient post. This protocol's structure to transmit a message is structured to exploit transitivity and historical encounter. It implies that the nodes have a repeated mode that is predictable but not a random movement. Such a move will result in the use of a fast and efficient post. Together, delivery predictability can be considered a probabilistic metric for an independently recognised destination. Each node calculates the probabilistic metric before sending the message, and the probability of node A meeting node B is the probabilistic metric $P_{(a,b)} \in (0, 1)$.

Moreover, predictability vectors are exchanged when two nodes meet each other. The carrier node passes the message to the neighbouring node when there is a high probability of the neighbouring node meeting the destination node. Otherwise, the message is not delivered further and kept to itself. Moreover, PRoPHET utilises a First In, First Out (FIFO) queuing or communicating mechanism for buffer management at the nodes. After forwarding to a minor number of full nodes, it results in the dropping of packets consistently. It is independent of any Messaging Scheme being used. Concerning node movements, slow and inefficient responses to evolving patterns could be carried out by this procedure. Besides, only the distribution predictability of nodes is used to determine the best carrier of the message later on PRoPHET. The PRoPHET used different parameters like Buffer, Location, Popularity, and Power to choose the best next hop to the destination. Some weight assigns these parameter values a well. To achieve an ultimate value from 0 to 1, these parameter values are normalized. A final metric is determined by the obtained values of these parameters and their comparable weights. This final metric then decides the best next hop. For selecting the next hop, such a complex protocol uses a lot of parameters. The calculations are difficult for these values, providing an outcome of various overheads in passing a message.

Discussing further, it does not use any parameters to predict the future actions in the movement of nodes in the network. Boldrini et al. [34] proposed the HiBOp (History-Based Routing Protocol for Opportunistic Networks).routing protocol based on opportunistic network history. To find a better path, this protocol uses the current history of the nodes, a screenshot of the area where the node currently resides (snapshot). The local context is stored in the Identity Table (IT) and the History Table (HT). They are shared, and the Current Context (CC) is generated when nodes meet one another. The HT previously stores the values that nodes see from the ITs. Any value is related to a Continuity Probability (Pc), Redundancy (R), and Heterogeneity (H) counter. All feasible background data is stored and used by HiBOp for a node. These are usually very difficult to discover. IT, CC, R, and HT need a significant measure of memory space on each node.

Essentially opportunistic networks are Delay Tolerant Networks (DTNs). There is no need for permanent source-destination communication for such networks; rather, attempt to utilise any current foundation to get the message over. For the transmission of the messages, hub adaptability is additionally figured out. The plan of any correspondence convention for DTNs emphatically relies upon how well the hidden portability is perceived [31, 35]. DTN conventions are created and tried in two essential manners: follow-based and versatility model-based [36]**.** On account of following-based engineering and approval, you can download a versatility follow from a confined number of following repositories [37]. These follow to get from this present reality and catch from the current versatility propensities for clients. Testing the convention on the tracks would give the following choice climate the most sensible presentation. Be that as it may, there are likewise a couple of disadvantages to utilising genuine follows a set number of follows, not all prospects being assembled, and not having the option, to sum up, discoveries based on a couple of follows. Due to these disadvantages, analysts have proposed models that catch key human versatility attributes and make engineered follows. Models are intended to impersonate hardly any real follow attributes, for example, between experience time because of the intricacy of comprehension and demonstrating human portability [38].

In many cases, a manufactured follow is approved by contrasting key attributes and the genuine follow. It doesn't assess the application-arranged boundaries of the created follow. Ongoing accentuation in DTN-related exploration has attracted another convention model that relies upon designs in human conduct. In these examinations, scientists endeavour to misuse the social parts of human portability to remove new projects and conventions [38]; for instance, analysts have assembled conduct arranged assistance called 21 profile-cast, which depends on client space-transient similarities [37, 39]. Profile-cast allows a deliberate system to utilise certain connections between versatile clients to advance and circulate intrigue-based messages in DTNs productively. Participatory sensing [23] gives a strategy to publicly supporting using versatile clients. This system empowers us to test and fabricate portability models while holding the relevance of the follows delivered. The intricate versatility model, the Time-Variant Community (TVC), is considered alongside the irregular course portability model as a contextual analysis in this work [40]. This model makes the non-homogeneous practices of portable clients in both existences. The following made by this model demonstrate (i) skewed visiting area preferences (ii) occasional time-subordinate return of versatile clients as found in WLAN estimations, alongside different measurements of experiences, for example, the normal level of hubs and meeting times. It utilises a few certified follows to affirm the nature of its design.

2.2 Similarity Analysis Patterns

Thakur, Helmy, and Hsu [41] discussed issues relating to mobile users' similarity, classification, study, and modelling. Using their online association matrix, they have adopted a behavioural profile based on user location preferences. The direct similarity method uses to define similarity and then calculate the behavioural gap to capture user similarity. This measures the distinction between the key Spatio-temporal behavioural patterns and can be used to organize users into similar groups or populations. Their results show a wealth of similar cultures of different consumer behavioural clusters in real mobile communities. This is true for all the traces tested, with the trend becoming constant over time. Surprisingly, however, the current mobility models do not capture similarities and lead to homogeneous consumers who are identical. Thus, the richness and diversity of customer behaviour are not captured to any degree in the current models. Such findings explicitly suggest that similarities should be captured explicitly in future mobility models. However, they do not present any solutions to the problems.

El Sherief et al. [42] discussed the concept of mobile user similarity as a primary enabler of innovative applications focused on opportunistic mobile encounters. It is analysed that the performance of established similarity metrics applies to our problem domain to quantify the fundamentally qualitative notion of similarity and propose a new temporal-based metric. Implement generalized profile structures, beyond mere position, in the form of a probability distribution aimed at capturing users' preferences and prior experiences. Afterwards, they Analysed existing and proposed similarity metrics for the proposed profile structures using publicly available data. Owing to the probability distribution structure of the suggested profiles, in addition to the classic Cosine similarity, as a good candidate for quantifying similarity, they define a distance metric from probability theory, namely the Hellinger distance. Besides, they implemented a novel temporal similarity metric based on matrix vectorization to capitalize on the richness in the temporal dimension and maintain low complexity. Eventually, the numerical results show several primary observations. Second, the temporal metrics generate because of the incorporation of dynamics into the temporal dimension on average, Relative to non-temporal ones, lower similarity indexes. Second, the Hellinger distance holds great promise for the quantification of similarity among probability distribution profiles. Third, vectorized metrics constitute a low-complexity technique. As a key enabler for innovative apps and services, they answered the question of mobile user similarity. To this end, they first introduced a new, generalized profile framework beyond mere location to capture the users' long-term interests and short-term experiences.

Lin et al. [43] looked at the possibility of finding a little privacy. Profiles are constructed as a way to help customers monitor their mobile app privacy preferences. Their word depended on the static code analysis to determine why the software needs and its permissions, distinguishing between. For example, applications that provide their core features based on specific permissions and applications need permission to share information with advertising networks or social networks. Under their latest solution, users will also be asked if the variance of forecasts associated with an entry in each profile will exceed a certain threshold. However, they did not introduce more advanced learning methods that could further improve the precision of such predictions.

For location-based social network services, mobile user similarity is essential. Research on measuring mobile device similarity based on their trajectories has gained much publicity with the omnipresence of location-acquisition techniques. However, Trajectories only indicate short-term mobile regularities, so it is difficult to capture the similarity of long-term user behaviour. Lv, Chen and Chen [44] discussed the long-term similarity of mining users' behaviours based on their trajectories. They suggested a two-stage approach to solve this problem. At the first point, the idea of the routine operation is proposed to capture the regularity of the long-term operation of users. A user's routine operations are taken from his/her daily trajectories. User similarity is determined hierarchically at the second stage based on the derived routine activities. Finally, based on both actual and artificial datasets, they tested their method. The experimental results show that, based on their similarity measure, users with different profiles can be discriminated against and thus show the efficacy of their method. Still, they have not proven that their method is percentage-wise better than other methods.

2.2.1 Mobility Modelling

For contact and knowledge sharing, cell phones are often used. A person usually carries a cell phone with him during the day. The data that a cell phone uses is another way of reflecting a cell user's mobility habits. It is, therefore, a valuable instrument for detecting patterns of human mobility. For this reason, data from cell phones is attracting researchers interested in researching patterns of human mobility. Modelling mobility offers a way of simulating mobility in ad-hoc wireless networks, urban planning, and emergency management systems. To explain the essence of mobility, mobility models use traces of mobility. Initially, human mobility modelling includes an appreciation of people's mobility preferences. Traces of mobility gathered through Global Positioning Systems (GPS) devices, social media accounts, and mobile phone networks can be used to deeply analysed human movements on a large scale. Besides, human mobility has a high degree of predictability, as most tend to spend most of their time in locations.[40].

Mobility patterning may usually be categorized as synthetic or trace-based. One of the most widely used synthetic models involves random models, e.g., Random Way Path (RWP) [45], where nodes are not associated and where their destinations are randomly selected in the simulation region. In several cases, the inadequacy of these models is well known in the literature [46] and will use it as a mere reference. Several recent models used trace analysis for design guidance. The truncated form of Levy walks observed in some human mobility scenarios is formed in [47] and is also used to recreate 22 distributions of power-law inter-contact time a Self-similar Least-Action Walk (SLAW) [15]. A time-variant population model using Spatio-temporal characteristics of human mobility was proposed by [48]. Munjal, Camp, and Navidi [49] research presented another model that borrows from SLAW and adds flight, touch, and pause time distributions is SMOOTH. Several other models have also Analysed the mobility patterns in greater depth, both motivated by the basic principle of human behaviour characterization, but they have not shown the experimental implementation of encounter prediction.

Our understanding of the connection between individual mobility and the social network is limited, partially because the collection of large-scale data is challenging, which at the same time capture dynamic traces of individual movements and social interactions. Due to many unknown variables that affect mobility patterns that range from the transit of jobs or constraints and priorities imposed by family, human trajectories are typically modelled using different diffusion models or random walks. Thus the modelling of human trajectories is done by the only sources, through continuous random walks with only small distances, usually between home and work or travelling for a short period. Thus, it is not apparent if the distribution represents individual users' movement or an undiscovered connection between population heterogeneities and individual human paths. However, this scenario changes quickly due to the widespread usage of mobile telephones. Indeed, the records of mobile communications gathered by telecom operators offer a significant representative of individual trajectories and social connections, tracking every telephone contact between the two parties and localizing the calling party place and time [50]. The increasing prevalence of mobile phones means that such data catches a significant part of a country's population. The availability of these enormous call-detail records has made it feasible to validate classic social network theories empirically. In fact, despite the inhomogeneous rates of the space and sampling, the vast amount of call detail record (CDR) data enables us to reconstruct many essential elements of daily routines, such as the most frequent visit places and their time and regularity. These data thus serve as a social microscope that helps examine mobility patterns together with social structure and their intensity [51].

From GPS in-car to smartphones and fitness wristbands, millions of personal gadgets link us to the cloud. These all-encompassing linkages between the physical and the digital environment provide many new possibilities for predictive mobility models. Each device user generates extensive information, which enables us to assess their regular mobility routines. This information is likely to affect a broad range, including health monitoring, computing, intelligent traffic management and catastrophe response. The modelling for mobility has evolved quickly over the last decade, but more work is required to address the individual predictability issue. Currently, mobile companies simultaneously gather complete data, including human movement and social contacts, across a broad section of the population. Almost every individual carry mobile phones throughout their everyday activities. When mobile carriers record the nearest mobile tower for billing reasons, each time the user uses his cellphone, the data is collected, and details of8 individual movements are recorded [52].

Opportunist networks consist of infrastructure-free networks consisting of wireless devices using their contacts for data transmission. Mobile devices don't depend on the complete availability of routes between source and destination nodes in opportunistic networks; instead, mobility is used to provide information. Mobile nodes use their neighbouring nodes to relay data so that the data reaches the target node. Since mobile nodes communicate data in opportunist networks when they touch one another, it helps to enhance routing protocol efficiency by anticipating the motion behaviour of mobile nodes. A rising interest in forecasting patterns of human life has led to the creation of many novel mobility models. It is interesting to model mobility since it predicts the pattern of interaction for a pair of mobile nodes. For instance, predict how long remains until a pair of mobile nodes meet each other again, given the movement patterns of mobile nodes.

Similarly, it is possible to forecast the length of contact between couples of mobile nodes and estimate the quantity of data transferable throughout the contact time. It is essential to develop models of mobility that can create synthetic trails to imitate actual human walking patterns. Mobility models may often be categorized into two models: synthetic and trace models [53].

The Random Waypoint Movement Model is the first simulation model used to measure randomly human mobility. Several random models of mobility have since been suggested. In general, synthetic mobility models are basic and straightforward to use; however, the traces of a synthetic model of mobility do not reflect human movement patterns. In particular, many research investigations have revealed that human beings do not move entirely randomly; instead, their movement patterns are somewhat predictable. There has thus been an increasing interest in creating models based on actual datasets. A mobility model that built real-life scenario data sets is termed a trace-based mobility model [54].

Point of interest is an essential task in location-based social networks that offer a customized recommended area for mobile users. Contrary to conventional product recommendations, POI recommendations are more complicated due to the time effects: checking whether the POI is suitable for a user's availability is necessary. Although many previous studies consider temporal impacts checking in timestamps just for modelling, they suffer from the scarcity of check-in data. In recent years, the introduction of positioning technology has collected a range of urban data about human movement.

Human movement patterns may be used from diverse sources of information to improve POI recommendations [55].

Gauss-Markov mobility model adjusts to various stages of randomness. There are fixed simulation regions in indoor mobility models in which go randomly. However, there is no notion of the simulation area in the outside mobility model as it is random. For low-density network situations, after that find the network connection under the Gauss Markov model is substantially lower than that achieved under the Random Waypoint model. The network connectivity achieved under the two mobility models is nearly similar for moderate and high-density network situations [56].

The long-term choices on the mobility of people rely on their present circumstances or future goals. Consequently, long-term mobility models should take into consideration lagging, contemporaneous and lead effects. Trajectory Distribution Approximation (TDA) solves the challenge of identifying new trajectory patterns in human and social networking. TDA has a comparable job to identify fundamental human movement patterns with conventional statistics-based scaling. It learns the individual trajectory probability distribution directly from the data. It is based not on a quantitative analysis of various factors and variables that affect user motion but on the underlying characteristic of its mobility patterns. The effective TDA solution should meet four desiderata: Ability to understand the natural features that determine moving patterns; allow a particular person to generate synthetic trajectories once the characteristics of their moving patterns are learned; ensure theoretical soundness of the way to guide and testing the results of the generation of trajectory; The synthetic trajectories should enhance the performance of other trajectory classification problems supervised, such as Trajectory-User Linking (TUL) in y Location-Based Social Networks (LBSNs) [57]. Neither the Generative Adversarial Nets (GAN), discriminators nor human labour can directly identify the actual semantics of produced trajectories. For example, the pictures or words produced may be readily distinguished from those realistic by manually identifying an ordinary individual without any supervision. When it comes to the point of intersections produced that form a path from actual places visited by a user, the final result may be identical patterns on a synthetic trajectory [58].

Previous work focuses primarily on simulating human movement spatial and temporal patterns. However, the semantics of trajectory is neglected and therefore, the motive of individuals behind the movement is not modelled. A new mobility model captures human movement using broad-scale semantic spatial-temporal data from social networks. This technique initially created a multimodal embedding method to leave users, times, locations, and activities unattended in the same place while maintaining the original trajectory. Then, the hidden Markov model was used to learn latent states and transitions between them inside the embedding space, which is the position vector to take spatial, temporal and user motivation into account [59].

Model associative prediction of position investigates a mix of data mining methods to deduce unchecked places secretly visited from the disclosed user route. Association rule mining extracts how often successive check-in pairs occur. The Markov model models trajectories with a series of observable and unobserved states or sites to discover the hidden sites of the user. The suggested ALPM thus employs a unified picture of the interaction between related check-in systems, people with similar trajectories and places located nearby [60].

2.2.2 Behavioural Characterization

Several studies have been proposed to use human behavioural patterns for intermittent linked networks. On this front, non-homogeneous behaviours in both space and time are commonly suggested by Daly, and Haahr [61], where the location of preferential visits and time-dependent mobility behaviour was studied. Chaintreau et al. [62] examined transfer possibilities. It showed that the distribution of intercontact time between wireless devices borne by humans is shown by a strong tail such as a power-law (that is, the time difference between two contacts of the same pair of devices). These properties have had a significant impact on the design and creation of new models of mobility that aim to capture realistic mobility [62]. Hsu, Dutta, and Helmy [63] identified Spatio-temporal similarity that can be exploited to propagate messages with similar characteristics through mobility models. Still, their focus is on a theoretical point of view, and they have not given any idea or model to save energy consumption in their work.

2.2.3 Co-occurrence Matrix

The use of co-occurrence data in spatial-metric research is very prevalent. For different purposes, data on co-occurrence can be used. For example, co-citation data can be used to study the relationships between authors or articles, co-authorship data can be used to study scientific collaboration, and word co-occurrence data can be used to create so-called co-word maps that provide a visual representation of the scientific field's structure. A transformation is typically applied to the data first when co-occurrence data is used. This transformation aims to extract data similarities or, more precisely, the normalization of results. For example, when Glänzel and Thijs [64] analysed relationships between authors based on co-citation data. To evaluate these similarities, they typically extracted similarities using multidimensional analytical techniques, such as multidimensional scaling and hierarchical clustering.

Similarly, when researchers use co-authorship data to investigate scientific collaboration, they usually apply data standardization and then focus their study on standardized data. In the research article by Jarneving [65], the focus is methodological, and there is no mathematical implementation concentrate on their work. They are researching different steps to draw parallels from data regarding co-occurrence. There are essentially two methods to extract similarities from co-occurrence results. Jarneving [65] referred to these methods; the techniques are sometimes referred to as the local and global approaches as the direct and indirect approaches. They are researching different steps to draw parallels from data regarding co-occurrence. There are essentially two methods to extract similarities from co-occurrence results. These methods are referred to as the direct and indirect approach, but the local and global approach is also referred to as the techniques. The indirect method of deriving similarities from data from co-occurrences is based on profiles of co-occurrences. A vector containing the number of co-occurrences between objects is a co-occurrence profile of an object. Indirect similarity tests measure the relationship between two objects by comparing the objects' co-occurrence profiles. The indirect approach is used primarily for deliberative data. The method is very well known from a theoretical point of view.

Two algorithms were used in a station's cellular system to predict the number system, namely PPM-C and LAST, by Monreale et al. [66]. It is considered that a heuristic model that predicts two factors is LAST. No user movement is one factor, and the other is the similarity of the former and next base stations. The move of market basket breakdown calculates the correct algorithm to be applied for a given portion of historical data. This movement takes place through the creation of diverse trends. Historical data from Market Basket analysis shows that the consumer stayed linked to a similar base station for four-time measures. Accordingly, the LAST algorithm estimated the new base station. In those cases where the historical data indicated a base station trend, there is another case where market basket research did not find a typical step. PPMC predictor was applied, and those algorithms returned the prediction on the single best position, which was the author's hard decision. The author uses the vectors for refunding all predictions, called the soft decision, and aggregates the hard decision with the soft decision. Still, none of these models focuses on reducing energy consumption during this whole process.

2.3 Encounter Prediction

Vu, Do, and Nahrsted [67] gave results of their locations, time prediction recently published in their research. The model consists of a time slot for days, weekends, and time slots, and the model based on the Naive Bayesian classifier was used. Bluetooth MAC addresses and locations listed every day is split into several records. The most probable (up to three) spot, mean calculation, standard deviation, and people (Bluetooth Max) mostly encountered at that location is predicted when the MAC address matches people encountered by the user looking forward to the form and time of time slot predictions. However, when they used the time slot, the result was not substantially modified, meaning that forecasts of the most commonplace, e.g. office, can be returned, and places that have been briefed for a limited time cannot be provided. Patterns of public travel, daytime, acreage use, and interest-point approach to a stable model. It can predict 60 % within an hour, depending on an individual’s next place. This work may be one of the first to add to the geographical models, and for position predictions, this geographical model movement pattern of the group to a person. Spatial-temporal data mining is applied to trajectories to predict the next area and a vehicle’s arrival time. But this accuracy is not enough.

Bauer et al. [68] proposed two methods, along with records, of potential forecasting destinations. A market basket survey of several buyers can establish the relationship between locations and the development of rules. For example, if anyone goes to the coffee shop, this algorithm decides the same direction the author shows on paper but has not demonstrated the required results if he or she should buy the cake or biscuits. In the cellular system of the base station, Lymberopoulos, Bamis, and Savvides did work. Two algorithms, namely PPM-C and LAST, were used to predict the number system. It is considered that a heuristic model that predicts two factors is LAST [69]. No user movement is one aspect, and the other is similar to the former and next base stations. The Market Basket breakdown movement calculates the correct algorithm for a given portion of historical data. This movement takes place through the creation of diverse trends. Historical data from Market Basket analysis shows that the consumer stayed linked to a similar base station for four-time measures. The most recent base station was predicted accordingly by the LAST algorithm. In cases where the historical data showed the base station trend, there is another case where Market Basket Analysis did not detect a typical movement at that time. The PPMC predictor has been introduced, and the algorithms return the prediction of the single best place, and this was the hard work of the author deciding. The author uses the vectors and aggregates the difficult decision with a gentle judgment [66].

2.3.1 Random Forest Model

The random forest model is based on decision-tree principles, where each tree stores and grows unique information in compliance with a particular parameter. In making the projections, the data is accumulated from the ensemble. Random trees from the forest produce fast performance, are simple to implement, and deal with numerous input variables without the overfitting problem. In this model, and tree in the Random forest is generated for each tree in the Random forest at the beginning. Input features of the node are selected and then further separated. The strongest one, based on the input features in the training dataset, the split is determined without pruning; the tree grows more similarly. The overall result is determined by taking the subtrees' average individual performances [70, 71].

Significant enhancements in arrangement and relapse precision have been demonstrated to be refined by utilising trees, where each tree in the gathering is created as per an arbitrary boundary. Extreme forecasts are reached by amassing over the troupe exhibited by Breiman [72, 73]. Since tree-organized indicators are the fundamental constituents of the group, and since every one of these trees is framed utilising a haphazardness infusion, these techniques are designated ‘irregular woods. The early work on mathematical element choice, the arbitrary subspace system, and the irregular split choice methodology decisively affected Breiman’s thoughts. Arbitrary timberlands have risen as genuine contenders to cutting-edge methods, such as boosting and supporting vector machines, as represented by different observational investigations. They are basic and simple to actualize; they produce exceptionally exact expectations and deal with countless information factors without overfitting. They are right now viewed as one of the most dependable broadly proper learning techniques accessible.

Genuer, Poggi, and Tuleau review [74] will offer each functional direction and a sensible beginning stage for understanding the strategy. Each tree in the assortment is underlying Breiman’s approach by first choosing a little gathering of info facilitates (additionally called highlights or factors beneath) at arbitrary, at every hub, to part on and, besides, by figuring the best part dependent on these highlights in the preparation set. Without pruning, the tree is developed to the most extreme size utilising the CART [79] procedure. Each time another individual tree is created, this subspace randomization strategy is mixed with stowing to resample and the preparation information assortment with substitution. Although the component shows up simple, it includes a variety of main thrusts that make investigating troublesome. Undoubtedly its numerical attributes remain generally obscure to date. Most hypothetical exploration so far has fixated on disconnected segments or adapted variations of the calculation. Fascinating endeavours with regards to this course are by Lin and Jeon [75], who make an association between irregular timberlands and closest versatile neighbouring techniques [62] for additional outcomes); Meinshausen [67] examines the precision of arbitrary woodlands in the feeling of restrictive quantile forecast; and Devroye et al.[39], Who give hypotheses of consistency for various streamlined renditions of arbitrary backwoods and other randomized outfit indicators. The ‘genuine’ arbitrary woodland measurable component isn’t yet totally known is as yet under dynamic examination. In the current paper, irregular woodlands by working out and cementing the properties of a model recommended by Breiman in [15]. Although this model is as basic as the ‘valid’ calculation, it is closer to reality than other plans. The short draft [15] is founded on instinct and numerical heuristics. Some are flawed and make the report hard to peruse and comprehend. In any case, the thoughts introduced by Breiman merit explaining and creating, and they will fill in as a beginning stage for our investigation troublesome.

Anyway, in the ongoing years, Artificial Intelligence (AI) methods have demonstrated power in stock anticipating. Numerous calculations, for example, SVM, ANN, and so on, have been read for heartiness in foreseeing financial exchange. Be that as it may, group learning techniques have stayed unexploited in this field. In [38] have utilised irregular woodland classifiers to assemble our prescient model, and our model has delivered truly noteworthy outcomes. The model ends up being truly strong in anticipating the future heading of stock development. The power of their model has been assessed by figuring different boundaries, for example, exactness, accuracy, review, and particularity. For all the datasets they have utilised, i.e., AAPL, MSFT, and Samsung, they had the option to accomplish precision in the reach 85-95% for long haul forecast. ROC bends were likewise plotted to assess their model. The bends graphically demonstrated the power of their model. It was likewise demonstrated that our arrangement calculation meets as more trees are added to the irregular woodland. The absence of standard information portrayals in EHRs limits the generalizability of prescient models. Candia et al. [38] proposed growing such models by first producing inferred factors that describe clinical aggregate. Decreases the number of factors, diminishes commotion, brings clinical information into the model structure, and modifies works away from the raw information portrayal, encouraging utilisation of standard information mining calculations. They consolidated this pre-handling venture with an arbitrary backwoods calculation to figure the hazard for readmission within 30 days for patients in ten malady classes. Results were promising for experiences that their analysis was appointed high or extremely generally safe. Relegating patients to these two dangerous gatherings could be an incentive to tolerant consideration groups planning to forestall readmissions. Computational models of clinic readmissions within 30-days can give essential data to recognise factors related to readmissions and distinguish patients at high and generally safe. Processing determined factors that speak to clinical aggregates and utilising them in hazard models may upgrade models' exactness and make them reusable in a more extensive assortment of clinical conditions. Clinical choice help instruments that influence these models in giving danger evaluations at the purpose of care could improve dynamic during hospitalization and the post-release period. However, they did not talk about energy utilisation and QoS in such a manner.

The exact energy expectation of the system plays an essential role in enhancing the energy proficiency of structures. A homogeneous troupe solution, i.e. the use of Random Forest

(RF) for hourly structure energy expectation, was suggested by Wang et al. [71]. A simple architecture of the Radom Forest Model, shown in Figure 2.1.

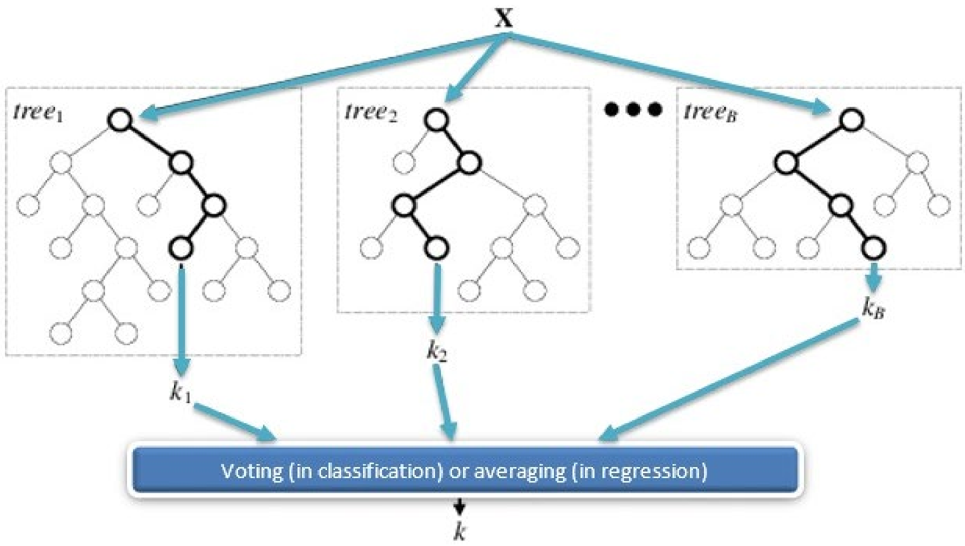


Figure 2.1 Architecture of Random Forest Model

The technique for estimating the hourly usage of the power of two instructive systems in North Central Florida was obtained. In contrast to examining the effect of boundary setting on the expectation execution of the model, the RF models prepared with different boundary settings were compared. The results showed that the number of factors (mtry) was not susceptible to RF, and the use of experimental mtry is optimal as it saves time and is more reliable. To approve the predominance of RF in building energy expectation, RF and relapse tree (RT) and support vector regression (SVR) were contrasted. The output list (PI) estimated RF expectation displays were 14-25% and 5.5 % better than RT and SVR, respectively, showing that RF was the best correlation forecast model. To identify the most convincing highlights during different semesters, an inquiry was also conducted based on the variable importance of RF. The findings have shown that the best highlights differ depending on the semester, indicating different operating conditions for the systems being attempted. A further connexion between RF prepared with annual and month-to-month information showed that the energy usage forecast for instructive structures could be improved by thinking about their energy behaviour adjustments across different periods. Table 2.1 shows the different model which used for prediction.

Table 2.1 Different Models used for prediction

Model use for prediction	Working/uses	Reviewed Literature
Random Forest Model	The most common algorithm, able to both classify and regress, is Random Forest. It can identify large quantities of data accurately.	[70, 71, 76, 77]
Gaussian Process	A non-parametric, Bayesian approach to regression that creates waves in the field of machine learning is Gaussian process regression (GPR). GPR has several advantages, working well on tiny datasets and measuring the predictions with uncertainty.	[16, 78]
Generalized Linear Model for Two Values	A more complex version of the General Linear Model is the Generalized Linear Model (GLM). Until drawing from an array of different distributions to find the 'best fit' model, it takes the above model to compare the effects of multiple variables on continuous variables.	[8]
Gradient Boosted Model	Until generalizing, the Gradient Boosted Model generates a prediction model consisting of an ensemble of decision trees (each one of them a' bad learner,' as was the case with Random Forest). As its name suggests, instead of the bagging used by Random Forest, it utilises the 'boosted' machine learning technique. It is utilised for the model of classification.	[79]
K-Means	K-means is a widely standard, high-speed algorithm that involves placing unlabelled data points based on similarities in separate classes. For the clustering model, this algorithm is used.	[80]
Prophet	In the time series and forecast models, the Prophet algorithm is used. It is an open-source algorithm created by Facebook that the company uses internally for forecasting.	[81]

2.4 Energy Consumption

One of the essential issues for the administration and organisation suppliers is the decrease of energy utilisation, which ought to be accomplished without effect on the nature of administrations. Tajiki et al. [82] proposed a novel asset distribution design that empowers energy-mindful SFC for SDN-based organisations. To this end, they model the issues of Virtual Network Functions (VNF) arrangement, distribution of VNFs to streams, and stream directing as improvement issues. Also, the model stream rerouting to decrease the effect of asset discontinuity on organisation usage. From that point, heuristic calculations are proposed for the diverse enhancement issues to discover close ideal arrangements on satisfactory occasions. The exhibitions of the proposed calculations are mathematically assessed over actual geography and different organisation traffic designs. The outcomes affirm that the proposed heuristic calculations give close ideal arrangements while their execution time is pertinent for genuine organisations. Their paper proposed a novel asset allotment design for VNF situations and directing in the Software defined networking (SDN) based organisations. Plus, administration work fastening to limit the energy utilisation and organisation reconfiguration overhead is tended to. To this end, it is a numerically defined issue and proposed a few heuristics to explain them. The goal is to limit the energy utilisation and organisation reconfiguration symptoms while the stream necessities are met. Recreation results demonstrated that the proposed plans dispense the organisation asset such that the energy utilisation is close to the ideal arrangement. Figure 2.2 shows the percentage of energy consumption in a different field.

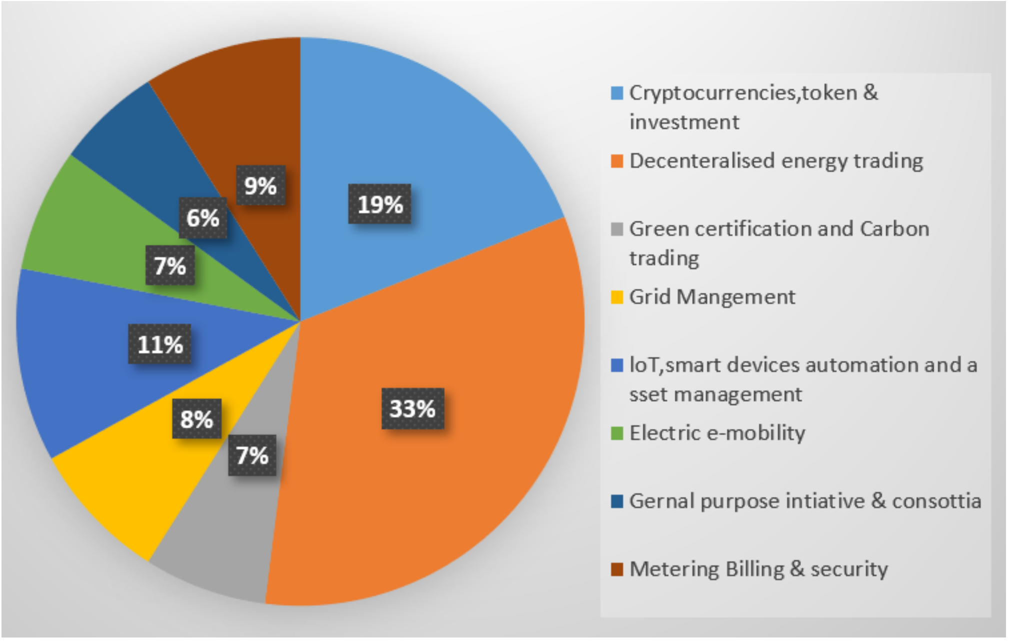


Figure 2.2 Percentage of Energy Consumption in different fields

Hsu, Dutta, and Helmy [69] suggested an addition to the epidemic. Routing Protocol by naming the protocol Spray and Wait. This protocol works by helping to minimise the distribution of the number of network-wide copies. Therefore, the congestion caused by the flooding is minimised. In-depth, the mechanism of forwarding has two The Spray Phase and Wait Phase phases are named. If a message spreads to L nodes only, L copies of messages are built or made. It is then referred to as the Spray Process. However, the relay nodes wait for the destination with a copy of the packet transmit it directly. The chosen value of L shows the likelihood of transmitting messages and is firmly responsible for this action. Again, L's calculated value is dependent on the parameters of the network. Even though L is chosen to reduce flooding, the protocol undergoes delays and resource usage issues. Makhlouta has presented an updated version of the spray-based routing scheme. This version of enhancement can categorize adaptive Fuzzy Spray and Wait.

2.4.1 Network Selection by taking Multiple-criteria decisions

The energy productivity of correspondence networks has been the focal point of wide-ranging exploration work in the past few years. In writing, several green methodologies were suggested to minimise the energy usage of both cell and fixed organisations, at all layers, and in all parts of the organisation (i.e. entry, metro/collection, and centre [40] investigated the effect that rest mode (SM)- put together green methodologies have on the dependability execution of optical and cell network components. First, they think about a gadget in disengagement (i.e. not connected to organisation inactivity), indicating how operational temperature Varieties of temperature and temperature, both presented by SM, affect its lifespan. At that point, they evaluate the impact of these lifetime varieties from an operational cost point of view, showing that a few gadgets are simple; for example, their reachable energy reserve funds probably won't spread the potential extra reparation costs coming about because of being placed in SM too habitually. They have researched the effect of SM-put together energy-productive methodologies for the lifetime of gadgets in cell and spine optical organisations. Specifically, they have first examined the issue from an operational cost point of view, demonstrating that the energy reserve funds presented by SM at times can scarcely take care of the reparation costs when the gadgets are supplanted as a result of disappointments.

2.4.2 Device to Device

The interactions between the fields of ICT and energy are becoming stronger and stronger. On the one hand, many ICT sectors are expected to play a key role in reducing energy consumption, especially those where most energy is spent, such as transport, construction, and manufacturing. On the other hand, with a fast growth rate, ICT is emerging as a greedy energy user. Marsan and Meo [83] briefly discussed three issues relating to green networking research's importance and potential impact, with particular focus on the wireless case, as this is the context in which energy efficiency is most needed. Their view is that economic incentive, at least in the first stage, would primarily stimulate research in the field of cellular networks. Although equipment manufacturers are working on creating energy-efficient components of the network (particularly base stations), the critical task of academic research is to develop algorithms to enable the portions of the network dynamically required for the provision of services to end-users with the desired QoS [75].

There is a growing concern among telecom operators about the energy efficiency of cellular networks. In addition to preserving their viability, the goal is also to reduce adverse environmental effects. Thus, the energy efficiency goal motivates standardization authorities and network operators to explore new technology to better network infrastructures. Since BSs consume a significant portion of the total energy consumed in a cellular system, turning off selected BSs (or placing them in sleep mode) when the traffic load becomes light is an effective technique to save energy on wireless networks. In the literature, BS sleeping technique has been thoroughly investigated. Switching off BSs results in substantial energy savings, as traffic in peak hours can be up to 10 times higher in the same region than traffic in off-peak periods. Yaacoub et al. [84] studied the joint operation of cooperative device-to-device (D2D) communications and explored green cellular communications. It describes a practical approach to grouping mobile terminals (MTs) into cooperative clusters. Through D2D communications, MTs cooperate in each cluster to share the content of shared interest. Besides, an energy-efficient technique is presented in an LTE cellular network to put BSs in sleep mode. Finally, to ensure green connectivity with both the users' MTs and the operator's BSs, both approaches are merged. D2D communications between mobile terminals have been investigated along with green cellular communication techniques in this paper. Collaborative D2D communications have achieved reduced energy usage for mobile terminals, while energy savings in the cellular network have been achieved by introducing the green communication strategy. Therefore, the combined approach outlined in this paper has contributed to energy savings for both mobile users and the mobile network operator.

2.4.3 Communication through the Internet

The web has modernized the contact of people connected eventually, likewise discover a few situations where the web foundation isn’t open for contact. Catastrophe zones, huge occasions, sports, and different celebrations, and territories where the web is either too costly or not even accessible are a portion of these zones. These pieces of the network need elective methods for correspondence. Delay Tolerant Networks (DTNs) model an organisation to settle shakiness in fixed information networks by connecting specialized portable gadgets. Backstrom et al. [85] introduced a nitty-gritty investigation of the as of late proposed impetus plans for Vehicle Delay-Tolerance Networks (VDTNs). This paper is composed to recommend and understand the recent improvements in the field of DTNs, zeroing in on the childish conduct of the hubs. Animating hubs in the investment organisation could improve its exhibition. All the hubs in the organisation make an honest effort to advance the messages to different hubs.

Be that as it may, the egotistical nature or the failure of a hub to submit the messages can corrupt the organisation's presentation. The childish conduct of hubs is a significant test in VDTNs. The hubs don't assume liability to advance the messages to spare its assets, for example, computational force, memory, and so forth. Different motivating force plans are proposed to address self-centeredness in the VDTNs to improve its presentation. These plans grant motivators to the hubs for dynamic cooperation in the organisation exercises. The paper gave some open difficulties. In the Internet of Things time, all over a trillion ordinary things will incorporate probably some capacity to store and handle data; moreover, and all the more critically, sharing that data over the worldwide Internet with the other trillion things. The mechanical objective is to incorporate the Internet and the web with everyday items (for example, entryways, seats, electric apparatuses, vehicles, and so on) and eventually interconnect the advanced and physical areas. The sorts of objects associated with the Internet, e.g. as far as to use, size and numbers, are incredibly assorted, consequently having distinctive calculation and correspondence requirements. Therefore, many processing structures and systems administration models have been proposed, and unique systems administration norms have been created. Khalil Massri and Alessandro Vernata [86] proposed a reference design for Delay-Tolerant Networking (DTN) steering conventions and a careful quantitative assessment of numerous patterns presented in writing.

Pandemic steering is a notable convention proposed to communicate a message to all hubs in an organisation [87]. A few ways, then again, abstain from broadcasting and depend on real properties examples of experience to control the message to the objective. In ongoing works, conventions have been proposed dependent on mutual activities in portable social orders, Simbet, Bubble rap, and Profile-cast, among others [61]. Messages are coordinated dependent on social gatherings and profiles in these conventions. These works utilise conduct attributes to support execution contrasted with directing conventions that are socially ignorant. Spyropoulos et al. recommended an expansion by naming it Spray and Wait for the convention to the Outbreak directing convention. By assisting with limiting the spread of the number of duplicates over the organisation, this convention works. The clog because of the reasons for flooding is additionally diminished. Top to bottom, there are two stages in the sending cycle, called the Spray Phase and the Wait Phase. If a message is just spreading to L hubs, for example, L duplicates of messages are manufactured or produced, it is called Spray Process. The transfer hubs with a duplicate of the message are trusting that the objective will send it legitimately. L's chosen estimation demonstrates the chance of conveyance of the message and is certainly answerable for this activity. Once more, the given estimate of L relies upon the boundaries of the organisation. Despite how L is picked to decrease flooding, the convention goes through issues, such as deferrals and asset utilisation. Be that as it may, the parcel of energy utilisation happened while utilising this one.

In [88], an improved rendition of shower-based directing plans was introduced by Makhlouta. This form of progress can be arranged by naming Adaptive Fuzzy Spray and Wait. This exploration aims to anticipate the comparative situation of walkers, and even specialists may amend the area update to shape a direction of movement. Regularly PDAs or gadgets with such helpless GPS inclusion would be killed, unremembered, or enter metropolitan ravines or other areas. For example, client warnings, such as substantial time slips between area changes, are primary realities that rely on such applications as Facebook and Twitter.

Moreover, clients must approach record cancellation or shirking of recording according to the alternative to approve complete protection strategies. Bauer and Deru recommended two different ways to anticipate future objections close to previous narratives [68]. A market container review of numerous purchasers is fit for recognizing the connection among areas and delivering rules, e.g. on the off chance that anybody goes to the café, this calculation chooses a similar heading the creator shows on paper, the calculation, yet has not demonstrated the necessary outcomes, on the off chance that the individual in question should purchase the cake or bread rolls.

2.4.4 Quality of Services Effects

In various ways, different cultures view and interpret QoS. For instance, it is referred to by network communities as the' measure of service quality offered to users by the network.' On the other hand, the Internet Engineering Task Force treats QoS as 'a set of service specifications that the network must meet when transporting a flow.' The primary aim is to provide QoS while optimising the utilisation of network resources. The network user community refers to QoS as the consistency experienced by applications/users. QoS is described by the International Telecommunication Union (ITU) as 'the capability of a

'Network or network portion to provide communications-related functionality between users. Quality services effects, shown in Figure 2.3.

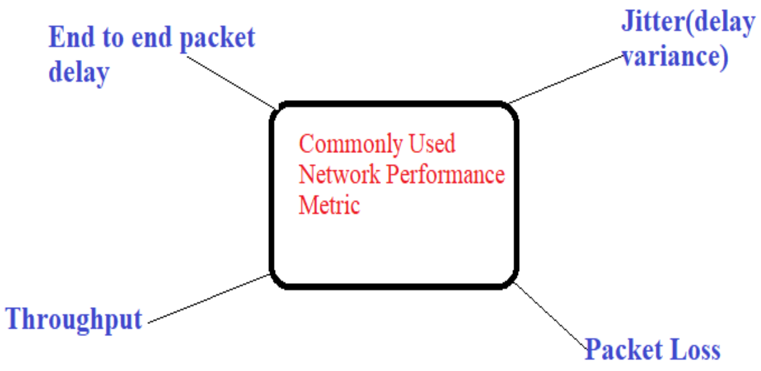


Figure 2.3 Quality of Services effects

The notion of QoS is quite large. QoS is more technically focused on IP networks, tracking and enhancing network efficiency metrics such as packet latency, jitter, loss of packets, and throughput.

Productive way calculation and supporting nature of administration necessities are key systems administration concerns. Current organisations have gone to multipath directing plans to part streams over various ways. Anyway, existing arrangements bring about sub-streams being coordinated in similar ways. QoS steering also presents significant difficulties because of the intricacy of multi-target way streamlining and association demands guaranteeing that all the necessities are satisfied. Programming characterized networks (SDN) have effectively conquered these restrictions with their brought together control of the organisation. Doshi and Kamdar proposed [89] a two-stage system for acquiring max-min QoS disjoint ways joined with the investigative chain of command measure in an SDN climate. The advancement and development of current organisations with SDN requires an exceptionally versatile and smart TE framework. This has advanced the objective of accomplishing a solitary physical organisation that supports changing QoS necessities and keeps up the proficiency of way improvement. Subsequently, in this paper, they talked about a two-stage technique that tends to these issues. The principal stage utilises the scientific progression cycle to catch the differing QoS necessities and reduce it to another cost work used to appoint connect loads. This lessens the unpredictability of preparing and improves the quality control of the organisation. The subsequent stage acquires k max-min QoS disjoint ways that make the heap adjusting of the central organisation proficient, lessens blockage, and improves unwavering quality.

With the consistently expanding number of versatile broadband information clients and data transmission concentrated administrations, the interest in radio assets has grown massively. One of the techniques utilised by versatile administrators to address this difficulty is to convey extra low-fuelled base stations (BSs) in territories of appeal. The following organisation alluded to the heterogeneous organisation (HetNet), which helps keep up the nature of administration (QoS) for a more significant number of clients by reusing the range. Yaacoub et al. [84] examined the issue of energy proficiency in heterogeneous cell organisations (HetNets) is researched utilising radio assets and force the board joined with the base station (BS) ON/OFF exchanging. The goal is to limit the all-out force utilisation of the organisation while fulfilling the nature of administration (QoS) necessities of each associated client. Coinciding microcell BS, little cell BSs, and secret femtocell passages (FAPs) are considered. In this paper, the energy productivity issue in thick cell HetNets is thought of. The specific instance of existing together microcell BSs, little cell BSs, and private FAPs are explored. In light of the support of private FAPs (missing, shut, half breed), three potential organisation situations are examined. A common technique for radio assets and force the executives and BS ON/OFF exchanging is utilised to use the radio access foundation and limit energy utilisation proficiently. The accessibility of privately produced environmentally friendly power is additionally consolidated in the created system.

2.5 Summary

The nearby devices will directly communicate with each other in D2D communications and establish a communication network. Data traffic can be offloaded to the D2D network instead of transmission via the infrastructure-based network. For example, by authorizing D2D communications, some users can download substances from the Cellular Base Station (BS), while others can get them from their associates. D2D communications and the energy use of networks [76] can also significantly minimise traffic congestion. Future wireless access networks can modify the user's BS affiliations and balance the traffic load between base stations [37]. One of their study topics is predicting consumer behaviour at the Massachusetts Institute of Technology (MIT) [90]. The time period is reported by mobile phones while linked to Cell Tower IDs and Bluetooth devices. And, depending on how the devices are attached, Bluetooth devices exhibit different behaviours. The participation of business students in a related area was found to carry out the same activities [91]. Bluetooth signals were mounted inside the person's house to check the accuracy of data transmission to locations via cell towers [92]. By predicting the next position of the goals using the dynamic Bayesian Network, an effective rate of 93 % to 99 % can be achieved. [93]. Zeibart et al. [94] predicted that driving to destinations will be feasible, given a partially travelled course, by comparing the probabilities of different possible routes. Along with past backgrounds, Lee and Cho [95] suggest ways of predicting future destinations. For the work, the Naive Bayesian classifier-based model was used in [96], consisting of the time slot of days, weekends, and 1-8 hours. A list of Bluetooth MAC addresses and locations for mobility prediction consists of the daily time stamp and is split into several records.

Chaper 3
Similarity Analysis

3.1 Introduction

The previous chapter presented a literature review on research on mobility prediction, user profiling, and encounters. The research studies reviewed have shown state-of-the-artwork to reduce energy resource consumption. This chapter's introduction starts with a brief detail of the mobility of the users, uses of such data, source for gathering human mobility data, mobility modelling, how it was carried out, and a few existing mobility modelling methods. After introducing human mobility, some common attributes of human mobility are listed: space-time associations of a user, periodical reappearance at specific locations, mobility patterns showing mobility preferences, uniform mobility through mobility correlations, device and network information, and mobility patterns other characteristics as well. After this, a brief introduction to the similarity in human mobility behaviour is provided; the next step is to identify the limitations of existing solutions for determining similarity in mobility patterns. The contribution of this research to the research field follows the boundaries of existing solutions.

The second section highlights the steps that are used in the proposed methodology for conducting a similarity analysis. A case study showing a paradigm for calculating similarity among different users using the current approach are discussed in detail. Some new characteristics of human mobility for their similarity analysis are selected. In the Similarity Analysis Approach (SAA) model, a cosine similarity-based mathematical model is presented in which the users' similarity is calculated, i.e. similarity in their mobility behaviour.

Section 3 provides detail of the experimental setup, the dataset used for experimentation, pre-processing of the dataset.

The results in section 4 showed the similarity of different users in the dataset at various locations. The similarity is then calculated on different days of the week, and a total similarity of all seven days is also shown in the result section.

3.1.1 Background

Researchers have employed various data resources, including banknotes, Tweets, and GPS devices[97]. This data can be helpful for traffic control [98], urban planning [4], and mobile marketing [68]. The data obtained from these sources can also help in empirically driven studies on human mobility. But these resources show limitations in scale, data resolution, and adaptation. Mobile phones have emerged as a more efficient resource for overcoming the constraints in behavioural analysis resources

The use of smartphones is increasing in people, becoming an essential part of our lives. The mobile phone is mainly used for communication and information exchange. An individual throughout the day typically carries a mobile phone with him. The data used by a mobile phone is another way of representing the mobility preferences of a mobile user. Hence, it is a valuable tool for identifying human mobility patterns.

For this reason, mobile phone data is attracting researchers who are interested in studying human mobility patterns. As with mobile data, many users can be tracked in the long term at a low cost. Some mobile users show similar mobility behaviour while others are dissimilar to one group but similar to others. This is called a similarity in mobility behaviour. Before discussing the similarity in mobility behaviours, first, the existing models for mobility modelling are focused.

Mobility modelling provides a medium for simulating mobility in ad-hoc wireless networks [35, 40] urban planning and disaster response systems. Mobility models use mobility traces to understand the nature of mobility. In mobility modelling, researchers usually collect the real-time data of mobile users carrying mobile phones with them during travelling, shopping, office, or university. The mobility traces are the result of human mobility. This mobility results from the repeated movement of a user to a specific place, time-based human activities, and mobility, because mobility is based on social relationships [4] The mobility of the mobile users may be at large scales, i.e. travelling from country-country, travelling through aeroplanes or trains. The mobility at a smaller scale can be travelling within a city, university, or from home to office. After collecting traces, the data is filtered out, and the daily life environment can be modelled.

Moreover, mobility patterns can be used to model people's decisions based on their behaviour. Through mobility modelling, various network operations, i.e. resource management, routing, handover, resource management, routing, and even a better independent deployment of connectivity models, can be simulated. The simulation is possible due to the possibility of understanding human movement patterns and regularities [4]

Modelling of human mobility, at first, understanding the mobility preferences of people is required. Mobility traces gathered through GPS devices, social media accounts, and mobile phone networks can be used to analyse human movements deeply at large scales. Moreover, there lies a high degree of predictability in human mobility as most of them tend to spend most of their time at specific locations. The time to visit a particular place is also mostly fixed [99].

3.1.2 Characteristics of Human Mobility Behaviour

Human mobility behaviour consists of different characteristics, specifically human mobility, location-time features, and pattern statistics. As most people are at common locations daily for a particular time interval, e.g. home/office, night/day, the mobility patterns and the time-space associations are similar at different scales. This similarity has been recently observed on real traces.

3.1.3 Spatial-Temporal Preferences

When different mobile nodes gather at a specific location, a wireless network is formed. The data generated by this network is usually community-contributed and mainly contains geographical coordinates, visiting preferences, social relationships, demographics, social-economic data, and many more. The in-depth analysis of this mobility data can help us determine a user's mobility preferences, favourite places for food, shopping, etc. The Heterogeneity of people or mobile users plays a crucial role in understanding and analysing human mobility patterns. Hence, the non-uniform behaviour of mobile users in space and time is captured by (i) Skewed location visiting preferences (ii) Periodical reappearances. The existing relevant literature shows that a mobile user exhibits mobility preference and reappearance at different locations and same time in DTNs. The idea of these mobility preferences and reappearances in space and time associations can make message distribution, delivery, and prediction of information transmission relatively easy and accurate in an opportunistic setting.

3.1.4 Skewed location visiting preferences

Recent studies on mobility traces show that most nodes do show the property of Skewed location visiting preferences. These preferences refer to visiting a small percentage of places for a small portion of the time. To create skewed location visiting preferences, the node in the network is randomly assigned to a sub-region of the mobility region R, called its community. A node tends to visit its community relatively more often than regions residing outside the community. The notion of community is different for different nodes.

3.1.5 Mobility Heterogeneity

During real mobility scenarios of human mobility, each node demonstrates a considerable pattern in its mobility pattern. These mobility preferences are quite different from those of other nodes. In some cases, few nodes show similar qualitative properties, but these properties often vary in location, access point, and stability time at the location.

3.1.6 Mobility Correlations

As evident, all nodes present in a network or other network cannot have the same mobility preferences or mobility characteristics. Instead, a subset of nodes will be subject to higher correlations in their preferred locations and visiting patterns. Consider, for example, colleagues or classmates, and they usually tend to be frequently present at the exact location or nearby each other because of their common study or working environment. These correlations are evident in the collected mobility traces. Generally, the larger the amount of average node mobility, the better the performance of the routing protocol that relies on such mobility.

3.1.7 IP address and MAC address

A user's mobility to different places causes it to connect to various networks. This connection with the networks assigns mobile devices to separate IP addresses. An IP address is a 32-bit long address assigned to each device when it connects to a network. An IP address can also provide geography information of a device [100]. An IP address has two parts: a network ID and a host ID. Network Id shows the information on the type of the specific network to which the device is connected. And the host id provides information about the particular host in the network. In comparison, the MAC address is a 48-bit unique identifier of each device.

Based on the above characteristics, different studies have modelled human mobility. These mobility traces of different individuals are highly correlated with their peers, colleagues, or other people [99]. The human mobility of different people shows similarity based on gender, educational background, or job. Multiple users periodically reappear at specific positions or locations [101, 102], creating a link among similar nodes [103]. People with identical mobility characteristics can show more similarities. Gathering various mobile nodes with similar mobility characteristics and patterns usually develops mobile societies in wireless networks [102, 104, 105]. Researches exist for measuring this similarity among different mobile nodes [106-108].

Hsu et al. [102] introduced a novel service that was behaviour-oriented 'profile cast, this service relied on the user and network coupling. It provided a framework to analyse and process the coupling of mobile users with the network for the transmission and propagation of data and information in DTNs. Some authors used mobile users' profiles (constructed using similarity data) for crowdsourcing. This Participatory sensing [30, 109, 110] was provided using recruiting campaigns and relied on similarity data of mobile users.
[106] analysed wireless users' behaviour using their traces from the network data. The authors used location-preference data for analysis and applied an unsupervised machine learning technique (clustering) to determine similarity in mobile users' behavioural patterns. The structure for data capturing was an association matrix. Based on the analysis's results, the authors suggested a 'distance' metric based on the similarity of the users.

In recent research, Kim et al. [111] demonstrated that if a user is connected to an access point (AP), the user keeps the connectivity unless the AP goes out of range and only then connects to the new best available AP. In Figure 3.1, a user is connected to a public AP, and even when the user arrives home, a better connection is available. The user stays connected to the less reliable/slow public network unless the user himself shifts to the better network or the public AP goes out of range. In a similar analogy, a set of two users, despite having a small Euclidean distance between them, might be connected to a different set of APs, and have different connectivity routes. This gives rise to the motivation that IP-based similarity has been used rather than Euclidean distance-based user's mobility similarity analysis.

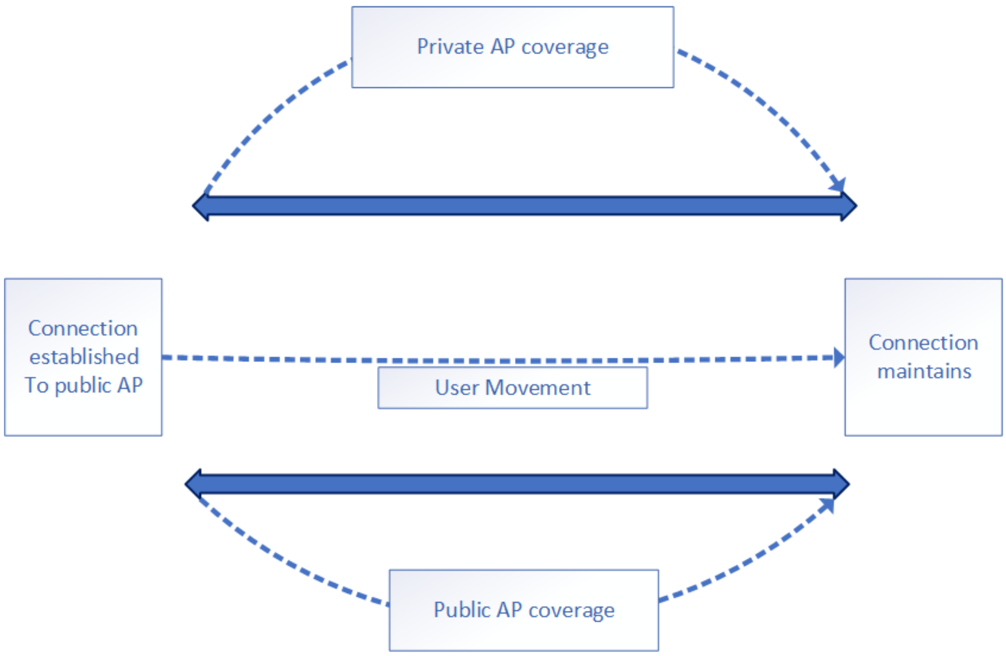


Figure 3.1 Mobile user connectivity behaviour with the Access Points

By keeping the scenario already mentioned above, in the presented research, the AP connectivity is kept as a sign of presence at a location, i.e. if two users are connected to the same AP, they are considered to be at the exact location. This increases the chance of 2 users having a direct connection in the MANETs scenario. This connectivity consideration helps us better correlate the user's mobility patterns and show an improved similarity.

In the SAA approach, the users' mobility has been mapped temporally, and the association has been made by comparing temporal presence at specific locations. The mobility characteristics of users, including temporal values, device information (IP, MAC address), and Access Point (AP) id, are considered. AP id is part of the IP address, hence simply and without loss of generality, the location using the IP address only can be referred to. This reduces the need to include AP as a feature in the data separately. The use of device and network information help to identify devices at slightly distant locations but connected with the same network. Network and device information reduce the resource consumption for searching a network for data transmission to those present in the same network. These characteristics also regarded users current at slightly distant locations as having similar mobility behaviour. Further, the temporal values allowed us to find similarities based on the tight coupling of the user, location, and time. The criteria for identifying users' similarity, analysis, and modelling based on their mobility patterns by employing a co-occurrence matrix on the selected characteristics are defined.

The purpose of this SAA approach is to identify the user's mobility characteristics that can efficiently contribute to determining the user's similarity without extra resource consumption. This research also aims to develop a technique for producing co-occurrences of mobile users based on their similar mobility behaviour. This similarity in mobility is identified through mobility patterns and can further determine mobile users' encounters in the next phases of research.

3.2 System Model for Similarity Analysis Approach

For determining users' similarity based on mobility behaviour, the IP address, MAC address of mobile devices, and temporal values (mapping of physical location vs time) are used. The IP and MAC address provides information about the access point to which the mobile is connected. The temporal values are used to estimate the exact time when the connection between the device and AP (each building has its access point) is established. These selected features are represented in the form of an occurrence matrix. The proposed approach considers users present on different building floors but connected to the same network with similar mobility behaviour. Hence, instead of using the various network for data transmission, the SAA approach uses the same network and reduces energy consumption by the resources used to search other networks. Figure 3.2 shows the overall flow of the working model that the Similarity Analysis Approach has followed throughout this methodology.

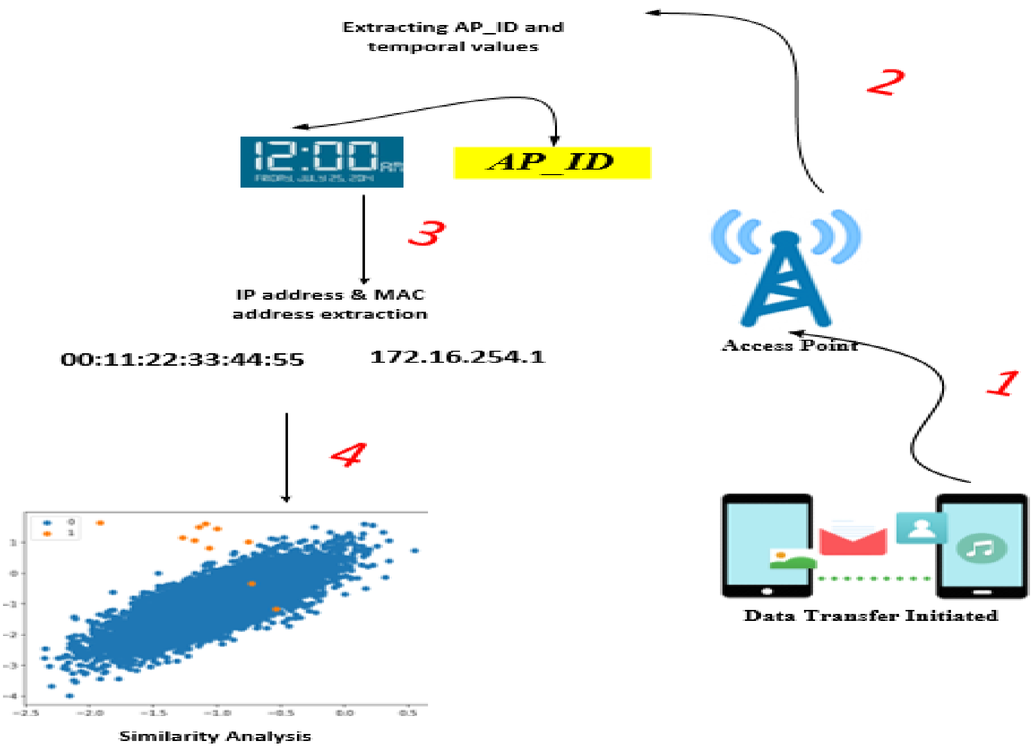


Figure 3.2 Working Flow of Similarity Analysis Approach

The SAA approach methodology is designed for scenarios requiring data transmission among nodes through a wireless network shown in point ‘1’. All nodes are usually connected to access points shows in point ‘2’.In point ‘3’ shows where the user’s AP_ID, device-network information, and spatial and temporal values are extracted. The extracted features are used in the SAA approach and computed the results.

3.2.1 Similarity Calculation:

The user’s similarity calculation consists of the following steps.

1. User’s location IP based location information at different times is represented as a matrix.
2. Capturing spatial-temporal values for occurrence matrix.
3. Similarity analysis with Co-occurrence matrix.

3.2.2 Information based on IP address

The proposed method model's the space-time associations of the mobility patterns and is thus better able to determine similarity among the users' mobility patterns.

A user's presence probability at a location is used to find co-occurrence with others as part of the model. The SAA approach used cosine similarity as a measure to estimate correlation among users. If the mobility patterns are correlated, i.e. mobility directions are the same, then cosine similarity enables the model to compute similarity patterns among the users. Here, it is essential to mention that the IP information is central to determining a user's geo-location. As previously mentioned, the IP address not only shows the device ID but network information as well. In the model, IP addresses have been used (both device no. and network ID), which resulted in constructing a more informed model. Hence better similarity patterns are obtained. The equation (3.4) similarity results are computed based on equation (3.1), equation (3.2), and equation (3.3). The model is explained in the next section, and its effectiveness is verified from the results presented later in this chapter.

3.2.3 Capturing AP and Temporal Values

The representation of IP address and time values is done in the form of an occurrence matrix. In this matrix, each row vector describes the normalised percentage of time the user spends at each location on a measured time, e.g. from 8 a.m to 8 p.m; in a day. This reflects the most significant locations of the user. Depending on the cases, the granularity of the time duration is defined as an hour, a day, or a week.

3.2.4 Similarity Analysis with Co-occurrence Matrix

This section explains some mathematical notations, where P is represented as an occurrence matrix with the order as $m \times n$. The 'n 'denotes Location (Each building has its access point), and 'm' represents a particular day. Let P_{uki} define the specific time in the k^{th} row at i_{tch} location of P. When M denotes the occurrence matrix of user 1,....n, is a matrix in the order of m x n. Let M_{aj} represent the time in i_t rows and J^{th} locations of M.

For i is not equal to j when M_{ij} equals the number of co-occurrence of time and location *I* and *j*.

$$M_{ij} = \sum_{k=1}^{m} p_{ki} p_{kj} \tag{3.1}$$

Suppose T_i denotes the total object occurrence or object co-occurrence at *i* location.

$$T_i = \sum_{k=1}^{m} p_{ki} \tag{3.2}$$

Suppose *Tj* denotes the total object occurrence or object co-occurrence at *j* location.

$$T_j = \sum_{k=1}^{m} p_{kj} \tag{3.3}$$

The F_c shows the ratio between the observed time of user *i* and *j* separately. The cosine angle is used between the i_{th} and j_{th} columns of the co-occurrence matrix of *P* for measure purposes.

$$F_C = \left(M_{ij}, T_{ij}\right) = \frac{M_{ij}}{\sqrt{T_{\mathrm{i}} T_j}} \tag{3.4}$$

In the above-explained model, the concept of cosine similarity is used to show the similarity of patterns in the mobility traces of the users. The user's movement pattern is considered a vector, and the angle between the vectors is inversely proportional to the similarity between the vectors, i.e. mobility traces of the users. The inherent idea of the cosine similarity is that the vectors are pointing in the same direction has a higher similarity irrespective of their physical distance. Here, if two users follow a similar movement path, they would have a higher similarity. The similarity is being modelled as a probability in this context.

3.2.5 Benchmark for Similarity Analysis

The paper similarity analysis uses the Ap and Mac addresses to build mobile user's spatial-temporal profiles[41]. The representation of spatial-temporal preferences in the form of an association matrix can be changed to use each column for a building (where a collection of access points represents a location/building floors and the time granularity can be changed to represent hourly, weekly or monthly behaviour. For this study, each row represents a day in the trace, and the column represents hourly presence at the location. After that, calculate the Singular Value Decomposition (SVD) of each associate matrix.

The SVD of a given matrix *A* can be represented as a product of three matrices: an orthogonal matrix *U*, a diagonal matrix *S*, and the transpose of an orthogonal matrix *V*. It is written as:

$$A = U * S * V^T \tag{3.5}$$

After representing a mobile encounter index metric, i.e. M_{ei} to indicate the similarity ranking (i.e. 0 represents dissimilar and 1 indicates the same) between two objectives' mobility preferences. Quantitively measure the similarity between them the eigenvectors as A =[a_1,a_2,a_{ra}] and B =[b_1,b_2,b_{rb}]. The mobility similarity can be calculated by the weighted sum of the pairwise inner product of their eigenvectors.

$$M_{ei}\ (A, B) = S_{A,B} = \sum_{i=1}^{rank(A)} \sum_{j=1}^{rank(B)} w_{a\ i} w_{b\ j} \left| a_i b_j \right| \tag{3.6}$$

M_{ei} (*A, B*) is a quantitative measure index that shows the closeness of two mobile peers in their mobility preferences of time-space dimension. *S* (*A, B*) is an index that shows the closeness of users in terms of the values in their co-occurrence matrix. Here, *A* and *B* are the respective eigen-vectors of the users. $w_{a\ i}$ and $w_{b\ j}$ are number of co-occurrences of user *A* and *B* at a specific location. Co-occurrence is measured in terms of time, location, and device information.

The similarity analysis shows that the value of the similarity is within the range of $0 \leq S_{A,B} \leq 1$. The lower the similarity rank indicates that peers have very different time-space mobility preferences. In other words, they have very less opportunities to meet each other somewhere. Otherwise, for the higher similarity rank cases, the user pairs have similar mobility preferences, or they have high probabilities to meet each other in some locations where they have the same interests to visit.

In addition, we can also quantify how the mobility preferences similarity between the same pair of moving objectives varies with time. Firstly, we can calculate the similarity between two mobile objectives e.g., A and B, at two points in time M_{ci} (A, B) T_1 and M_{ci} (A, B)T_2, where T represents a specific time instance. We perform this calculation to all user pairs and investigate the relationship between the users' location and time we calculated the mobility correlation coefficient M_{cc} of the similarity matrices obtained after a T interval as; They also quantify how the mobility preferences similarity between the same pair of moving objectives varies with time. First, the similarity between two mobile users is calculated, e.g. A and B, at two points in time M_{ei} (A, B)T_1 and M_{ei} (A, B)T_2, where T represents a specific time instance. This calculation is performed on all user pairs and then calculate the mobility correlation coefficient (M_{CC}) of the similarity matrices obtained after a T interval as:

$$M_{CC} = \frac{\sum_{A,B}(X-\bar{X})(Y-\bar{Y})}{NS_X S_Y} \tag{3.7}$$

where X = $S_{A,B}(T_1)$ and Y = $S_{A,B}(T_2)$ and the notations $\bar{X}$ and S_X Denote the average and standard deviation of X, respectively. N is the total number of mobile user pairs. This mobility correlation coefficient indicates how stable the relationship between the mobile objective pairs is. It has been reported that the similarity metrics between objective pairs correlate reasonably well if the considered time is not far apart.

3.3 Case Study

This section contains a scenario for understanding the proposed model and the benchmark working in detail. A case study shows that the data transmission distance is a key but changeable factor contributing to the overall energy consumption. It is straightforward that the shorter the transmission distance, the total energy consumption will be reduced. For example, there are two users Alice and Bob: Alice and Bob are working in the same university but on a different floor, Alice is a professor at the university, and bob is studying, and they usually met to discuss the research work. Alice intends to send a file to bob, which will be required by bob after seven days as at the moment, Bob is working on another task. In this case, how can Alice send the file to Bob? There are several options available:

Option A: Alice can send the data through the internet, passing through many intermediate devices and covering several hops.

Option B: when a direct file-transfer service is available, for example, they both have their routers, thus can exchange data between each other through point-to-point transfer.

Option C: Check similarity among the users' mobility traces if found a similar option for a direct file transfer.

The first two options, A and B, are the most popular way and overall energy consumption depends upon the size and distance of data.

Option C, loose requirements of time, and a delay-tolerant indicator is decided to reach data between that duration. In option C, human mobility traces are matched according to behaviour. Alice and Bob have similar mobility preferences as they are at slightly distant locations. They are also connected to the same network. In such a way, the proposed approach can fully utilise the benefits of mobility traces, using these traces, and predicting the similarity of location and time. If they match, then data will transfer by D2D while consuming less energy.

The benchmarked work has calculated the similarity of the users based on the use of their Spatial-temporal preferences and preferential attachment to locations and the frequency and duration of visiting these locations. They have utilised an association matrix that captures Spatial-temporal preferences and a statistical technique that use it to measure similarity among mobile users. In the proposed architecture, an algorithm based on a co-occurrence matrix that calculates the similarity of the users based on their occurrence to visit a specific place is developed. For comparisons, the same case study for both the benchmarked and SAA approach is utilised.

3.3.1 Calculation of Similarity Analysis Approach

The case study is designed as four people moving to specific locations. The data window has been set to 4 days, i.e. 4 days of the mobility data have been shown here to demonstrate the working of the SAA approach. So, assume that there is 4 users matrix, and each matrix represents the location (Column) and day/time (row). The location has been labelled based on the IP address of the user. The IP/locations of the users have been numbered for a more straightforward representation. The list of IP addresses have been mentioned against the location numbering assigned linearly:

1. 138.70.128.43
2. 168.233.57.204
3. 95.36.236.124
4. 194.134.255.229
5. 51.136.182.224
6. 99.245.73.36
7. 98.201.125.212
8. 202.189.184.98
9. 56.76.244.60
10. 37.62.113.116

Each location number in Table 3-1 represents an IP address of the user present at a particular location. The data in Table 3-1 has been obtained after cleaning and pre-processing the mobility traces logs of the users. Here, for each user, his/her presence at a location is noted in the cells of the following matrices. Each cell in the following tables represents an event in the daily mobility history of the user. Considering the example of user 1, the first cell shows that user 1 (identified through the MAC address, xx:xx:xx:ed:eb:73) was present on location (based on IP address) at the given time t. Table 3-1 matrices of 4 users are generated as a sample dataset.

Table 3.1 The matrices of 4 users

		User 1 (xx:xx:xx:ed:eb:73) mobility history						User 2 (xx:xx:xx:a7:c0:c2) mobility history			
		Time						Time			
		t_1	t_2	t_3	t_4			t_1	t_2	t_3	t_4
Day	d_1	1	3	1	5	Day	d_1	1	1	2	3
	d_2	2	1	3	4		d_2	2	1	3	4
	d_3	1	4	2	3		d_3	3	4	2	1
	d_4	2	6	1	2		d_4	4	6	3	6

		User 3 (xx:xx:xx:96:65:93) mobility history						User 4 (xx:xx:xx:56:dc:a1) mobility history			
		Time						Time			
		t_1	t_2	t_3	t_4			t_1	t_2	t_3	t_4
Day	d_1	1	2	6	4	Day	d_1	2	5	1	3
	d_2	1	2	2	1		d_2	4	1	7	9
	d_3	2	7	1	2		d_3	6	2	3	2
	d_4	1	2	3	3		d_4	7	2	1	4

The day number in the above data represents a date from the mobility traces of the users, and each day represents a date as listed in Table 3-2

Table 3.2 Each day is representing a date

		Date
Day	d_1	Jan 23
	d_2	Jan 24
	d_3	Jan 25
	d_4	Jan 26

The time number from t_1 through t_4 represents the hour number in the day. Here in the sample example, the data of 4 users, 8 locations, and 4 hours of each day, i.e. 8 a.m., 9 a.m., 4 p.m., 5 p.m., is shown in Figure 3.2.

After the representation of the data in the matrix form, the analysis of the data is accomplished. Here, only the data of 4 users are represented. The proposed method of similarity has been explained through equations (3.1) to (3.2)

After generating the 4 user matrices, the SAA approach needs to make the user's pairs and calculate the cosine similarity with equation (3.4). Each cell in the cosine similarity matrix represents the similarity between two user's mobility for a day. The results of the applying equation (3.1) to the data of user 1 and user 2 is given in Table 3-3.

Table 3.3 The data of user 1 and user 2

		User 1 (Day)			
		d_1	d_2	d_3	d_4
User 2 (Day)	d_1	21	28	22	55
	d_2	21	30	20	47
	d_3	18	24	26	52
	d_4	16	21	34	59

The result of applying equation (3.1) at the user 1 and user 2 data. The above table simply shows the number of co-occurrences between the users.

After calculating the co-occurrence and magnitude of the angle between the eigenvectors of the user's daily mobility data, the similarity between them is calculated given in Table 3-4.

Table 3.4 The co-occurrences between the users

		User 1 (Day)			
		d_1	d_2	d_3	d_4
User 2 (Day)	d_1	0.904	0.852	0.669	0.931
	d_2	0.990	1	0.667	0.871
	d_3	0.849	0.800	0.867	0.964
	d_4	0.616	0.572	0.925	0.893

It can be observed from the user 1 and user 2 data given above that they both follow the same route throughout day 2. They have a similarity of 1 for d_2 of the observed routes.

Table 3.5 shows cosine similarity among the mobility patterns of the users for each specific day. It is essential to mention here that the similarity analysis is not symmetric instead of the general notion of having a symmetric matrix of cosine similarity. In some of the applications where similarity is used, the diagonal of the matrix is 1. In the SAA approach case, the matrix is not symmetric. It is because the rows and columns are the mobility patterns of the different users. If the same user is selected to calculate the similarity with him/herself, then the SAA approach returns a symmetric matrix.

Table 3.5 The similarity between different days of the same user

		User 1 (Day)			
		d_1	d_2	d_3	d_4
User 2 (Day)	d_1	1	0.84	0.90	0.86
	d_2	0.84	1	0.66	0.74
	d_3	0.90	0.66	1	0.88
	d_4	0.86	0.74	0.88	1

The application of equation (3.4) is to find out the similarity between different days of the user days, as shown above. The cosine similarity shows the angle between the eigenvector of the mobility traces of the users. Each cell of the table above represents the similarity in the mobility pattern of users for given days, for example, considering checking the similarity between 2 users for a given day. The above matrices are filled just to show the calculation. Otherwise, cells at the diagonal show actionable similarity, as it can be used to establish the 2 users could be present at the same location at a point in time. This similarity analysis is essential for data transmission between the users.

3.3.2 Calculation for the Benchmarked

The same matrix is used in the SAA approach and the benchmark. In the benchmark, SVD is calculated for each matrix. SDV (Singular Value Decomposition) converts high dimensional data into lower dimensions to handle large datasets denoted by A.

$$A = U * S * V^I \tag{3.8}$$

Where U is t × t matrix, S is t × n matrix, V is n × n matrix. It is being calculated transpose of A and with its eigenvectors. The similarity for each pair needs to be calculated. The similarity is to be calculated as the weighted sum of the inner products of two users.

The similarity analysis is conducted as:

(User2, User1), (User4, User1), (User3, User2), (User4, User2), (User4, User3)

For example, Similarity values for each pair are shown in Table 3-6.

Table 3.6 Similarity of two users

	u1	u2	u3	u4
u1	0.721926	0	0	0
u2	0	0.28716	0	0
u3	0	0	0.951263	0
u4	0	0	0	0.974081

After calculating the similarity of each pair, the Weighted Similarity is calculated in Table 3-7 for comparing the SAA approach with the benchmark similarity results.

Table 3.7 Benchmarked Similarity Results

	u1	u2	u3	u4
u1				
u2	0.733607			
u3	0.733607	0.66279		
u4	0.320123	0.673118	0.855155	

While SAA approach Similarity results are given in Table 3.8.

Table 3.8 Similarity Analysis Approach Similarity Results

	u1	u2	u3	u4
u1				
u2	0.7761			
u3	0.7861	0.659836		
u4	0.9336	0.820487	0.875770883	

After analysing both baseline and the proposed results, it is observed that the similar pattern of the SAA approach is better than the benchmark. It shows that matrix information and the direct similarity method used in the SAA approach are advanced than the benchmark method.

The model calculation shows the direct similarity measures and compares their properties with the benchmark. The similarity is calculated through the direct similarity method; it applies to co-occurrence data. The purpose of the direct similarity method is to normalise the data. The direct similarity method corrects the data for differences in the total number of occurrences or co-occurrence of objects.

3.4 Numerical Studies

In the numerical studies, all data description, location information, and how the data is gathered are given and explained that how results are generated based on the data.

3.4.1 Dataset Description

This section provides the details of the traces used for the experiments. The traces are actual measurements taken from the University of Southern California USC. It's a very recognized dataset and is used in different research works. The data contains information about WIFI associations and user profiles. The data is about different buildings on the campus, access points, and data about other days, including weekdays and weekends. The dataset is about the campus of USC in Fall 2007. The mobility traces dataset of the campus that is used is shown in Table 3.9.

Table 3.9 Mobility traces datasets

Campus	# Users	Duration
USC	3000	Fall 2007

Table 3.9 provides the detail of sources for these documented network usage traces. The campus environment is chosen here since it is comprehensive` with many active users and has sufficient location samples. The data consists of Wi-Fi usage and location data.

3.4.2 Location Information

To analyse traces in different places, location information of other Access Points is required. Since exact locations of the APs were available, the APs are assigned approximate locations based on the campus building where the AP's were installed.

3.4.3 Data Gathering

The event logs of the WLAN event is collected from the university campus. Every log entry had a timestamp, an IP address at a corresponding access point (AP), and the MAC address of the associated user device. There are almost 1,700 Access Points and $\approx 78k$ devices found in the dataset. Therefore, in this research, the behaviours of devices were analysed as indicated by their respective MAC address. The Paired user mobility patterns of behaviour are represented by the patterns between mobile nodes in different users. When two user devices are associated similar to Access Point where there time in overlapping intervals. Traces are then created basing Wireless LAN Access Network (WLAN). The location information of Access points is needed for the analysis of traces in different places. Since the APs exact locations are not available, the APs are given new proximate locations based on the building where they are installed. The crowdsourced service validated positioning from 130 matched Access Points of about 7.6% in around 58 building structures within a range of 200m or less from where they are mapped. Almost 1.5% of campus area error was considered a reasonable maximum Access Point range of coverage, coarse-grained localisation that is inaccurate services. The coordinates of the centre of each building are used in which users may see an AP of the building.

3.5 Simulation Studies

At first, from the dataset, the SAA approach abstracted out a subset of peers independent of each other as much as possible from the substantial population data. Then the relevant statistical analysis is conducted on mobile peers' selected co-occurrence matrix information. This matrix consists of information about the AP, temporal values, IP address, and MAC address of the mobile devices. The dataset was divided based on each day, i.e. for seven days a week. Each day, the iterator tools are used to bring the data in sorted form, i.e. all combinations were extracted. After that, the direct similarity technique is used to normalise the extract of a mobile user's preferred locations and times. The locations visited by both users x, y were sorted, and their co-occurrences were recorded. After recording multiple co-occurrences, the redundant information is removed and unique location IDs were extracted. The result was a co-occurrence matrix having no null value and combinations of visited locations for both users. Each similarity value was calculated for different days and plotted respectively. An Instance of the co-occurrence matrix in this work is shown in Figure 3.3.

	Time	IP_Address	Some_value	MAC_Address
0	Dec 23 08:13:00	128.125.214.245	6	xx:xx:xx:ed:eb:73
1	Dec 23 08:14:38	128.125.214.242	5	xx:xx:xx:b1:dc:90
2	Dec 23 08:41:47	128.125.214.241	5	xx:xx:xx:a7:c0:c2
3	Dec 23 09:39:36	128.125.214.243	3	xx:xx:xx:1:6b:4a
4	Dec 23 09:48:32	128.125.214.186	1	xx:xx:xx:6:95:fc
...	...	...	...	...
1128839	Sep 27 16:51:03	128.125.214.245	2	xx:xx:xx:81:77:bc
1128840	Sep 27 16:51:05	128.125.214.244	4	xx:xx:xx:81:77:bc

Figure 3.3 An instance of the output co-occurrence matrix

3.5.1 Simulation Results

Based on the analysis of similarity profiles, the following results are found. At first, different user pairs were shown that were at a specific location at the same period.

For calculating mobility similarity among other user pairs at different time intervals, data of the users are selected from the USC traces available in the dataset. Figure 3.4 shows that the other pairs of users are calculated at the exact location at the same time interval.

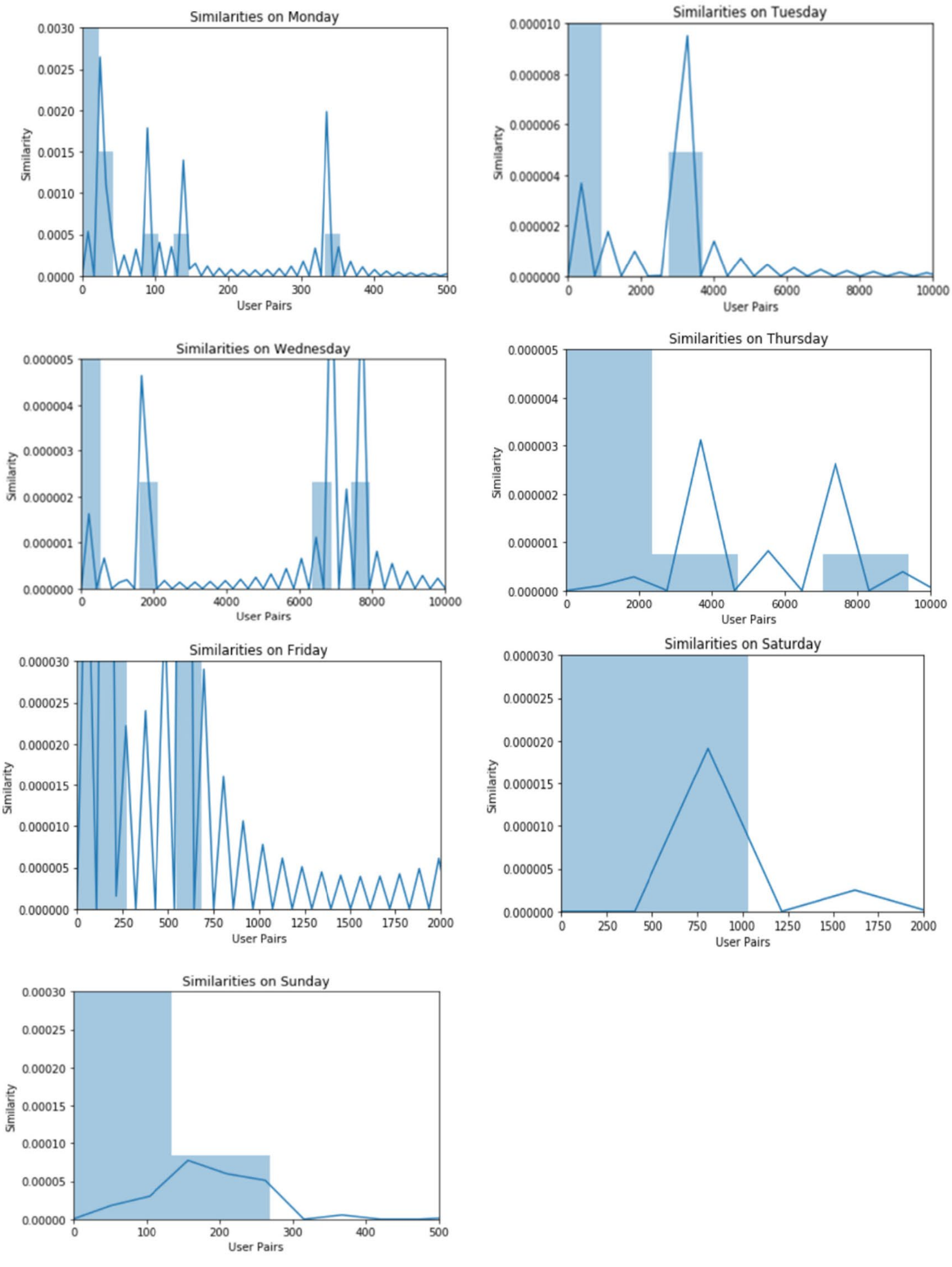


Figure 3.4 Similarities of user pairs on different days.

Following histograms show the similarity of different user pairs on additional days of the week. The users are from the same dataset that is used for experiments. The number of user pairs is considered a similarity function in these histograms, showing different mobile users' behavioural similarities.

From the histograms in Figure 3.4, most of the users showed high behavioural similarity on Wednesday. Friday also observes the highest similarity but with less number of user pairs compared to Wednesday. The highest similarity on Friday is 0.000030. The lowest similarity on Friday is 0.000005. On the weekend, i.e. Saturday and Sunday, the lowest similarity reaches zero. The highest similarity is 0.000020 and 0.00008. The number of active users is relatively less due to the absence of any academic activity these days, the similarity of all days in the week.

Figure 3.5 shows the overall similarity of all days in the week. Each curved line represents the similarity of user pairs on a specific day. As described earlier, Saturday and Sunday show similarity almost approaching zero. At the same time, Wednesday and Thursday have the highest similarity of the user pairs.

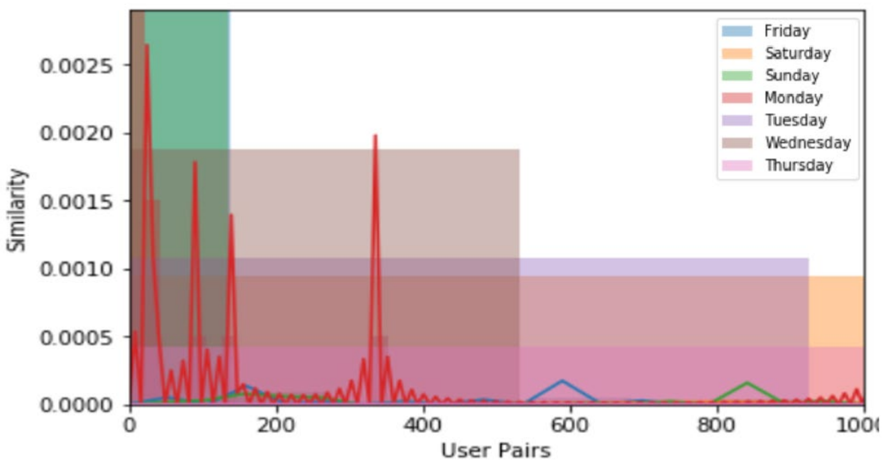


Figure 3.5 Similarities of user pairs

The comparison results show a similar pattern with an increased number of similar user pairs in the SAA approach, Figure 3.6, the x-axis shows similarity, and the y-axis shows user pair. This plot shows the comparison of the SAA approach with the existing work. This is very clear that the existing model can predict the similarity of 6000 users. Still, the SAA approach can indicate the similarity of 9000 users, so it is clear from the results that the SAA approach is 33 % more efficient than the existing approaches. Comparison results of existing and both approaches based on same dataset. In the proposed work, user pairs vs similarity are higher in comparison with existing work.

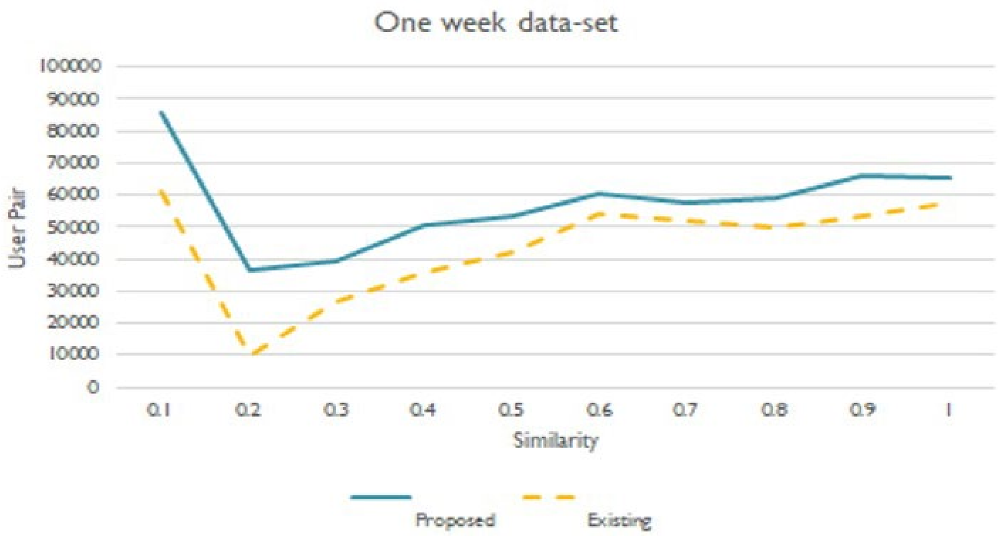


Figure 3.6 Similarity Analysis Approach similarity comparison with existing work.

Comparison results of current datasets and both processes based on one week of datasets. Yellow(dotted line) shows the existing model results, while blue(straight-line) shows the SAA approach (proposed model) results.

3.6 Summary

This chapter covers the method for finding behaviour similarities of mobile users. In the first stage, how our mobile phone can be used as a user's mobility behaviour is analysed listed. As most people carry their mobile phones with them throughout the day, it is the most suitable and cost-efficient way to analyse mobility behaviour. The existing mobility modelling methods are also listed. For data transmission among users, there is a need to find the similarity among user's behaviour. The SAA approach has used spatial-temporal parameters of a mobile user and the device-network information of the mobile node. The spatial preferences of a user and temporal values at which a particular location was visited are listed. The mathematical way of calculating the similarity was also provided in this chapter. For experiments, the SAA approach has used the mobility traces of students from a university campus of USC. After extracting the spatial-temporal associations of traces, the SAA approach stored them in a co-occurrence matrix. After that, the SAA approach normalised the time associations and applied a direct similarity technique to normalise the data and extract the preferred locations and times of a mobile user. Through the case studies, the whole process of calculating behavioural similarity is explained. In experiments, the similarity analysis of the processed data is performed and calculated for different users on different days. In the next chapter, the SAA approach will be enhanced with encounter prediction.

Chaper 4
Encounter Prediction Using Random Forest Model

4.1 Introduction

In chapter three, work for calculating similarity in mobility behaviour of different mobile users was presented. The similarity was calculated based on their familiar mobility preferences, i.e. visited locations and the time of visit. The co-occurrence of the Spatial-temporal traces has been computed, and similarity analysis of the users has been done. The experiments showed that a higher accuracy could be achieved than the existing work [112] for the estimation of similarity among a large number of users. After calculating the similarity, the next step is to predict the encounters of different users. When and where users might come into proximity based on their common mobility preferences. This chapter works on encounter prediction, where encounters among different users are predicted based on their previous mobility history. An encounter is predicted in the case of matched mobility preferences, i.e. same time to visit a location. Predicting future encounters helps avoid additional resource consumption for continuously predicting future locations and finding a meetup point and time for initiating data transfer.

At first, the introduction and the need for encounter prediction are mentioned. Then some existing works and techniques for encounter prediction are listed. The presented model used for encounter prediction is explained along with the mathematical representation of the model. The experiments, results, and comparison with the benchmark technique are also discussed in detail in this chapter.

In the current revolutionized digital world, smartphone users are increasing exponentially [113]. The proliferation of these devices results in increased mobile traffic, which has reached up to 79% in 2019 [114]. The mobile devices form a wireless Mobile Ad-hoc Network (MANET), which provides different scalability and flexibility advantages. Other users can join or leave the network without any prior notification [81]. The users of these mobile devices travel to different places autonomously in any direction. This mobility of the nodes enables them to dynamically and arbitrarily connect with other such nodes. The mobility of users makes the MANET environment unpredictable for real-time applications of data transfer [115].

Mobile ad-hoc networks do not require any fixed infrastructure [116]. Due to the absence of any central infrastructure and limited transmission distances, the nodes in MANET are responsible for discovering each other for data transmission [116]. Discovering other nodes is also essential for reliable communication and ensuring data transmission without interruptions [81]. The forecasting of the following location of a mobile user is called mobility prediction. For mobility prediction of a specific mobile user, where millions of devices are in the mobile network, determining the behavioural pattern of users is a key step. As most individuals follow a standard pattern, for example, a user is at work from 9 am to 5 pm, goes back home at around 5 to 6 pm, and is usually found at a restaurant between 7 to 9 pm. From these behavioural patterns, it is evident that the users visit certain places around specific time intervals. Using such mobility patterns, the future location of a mobile user can be determined. This pre-determined future location can lead to encounter information among users. The time and location information of mobile users is important information in this scenario. The mobility of mobile users is quite homogenous and usually shows periodicity. The presence of two users at the same place at a specific point in time is being called an encounter. For a particular encounter to occur, a different set of users, $u_1, u_2, \ldots u_n,$ must be present at location l and at time t.

There exist several works for future location prediction, but only a few focused-on encounter prediction. Hui and Crowcroft [117] showed the mobility prediction of users, they considered the idea of centrality, i.e. at first, they predicted the nodes having centrality values and considered them as hubs. Based on this idea, a message packets forwarding algorithm was proposed. For evaluating their approach, the authors used a greedy algorithm, 'RANK' and observed the message delivery ratio and cost of their proposed algorithm. Further work by Hui, Crowcroft, and Yoneki [118] proposed 'BUBBLE Rap' hybrid algorithm, that uses social data about the nodes to calculate the probability that a neighbour is quite near to deliver a packet. Ciobanu and Dobre predicted future device encounters by employing a function that describes individual mobility patterns, e.g. the Poisson distribution [119].

There exists a considerable amount of work on future location prediction. In most of the studies, after discovering a future location, the sender node ensures the occurrence of the encounter. These operations of mobility prediction, encounter checking, and data transfer across multiple nodes result in massive resource consumption, such as bandwidth, energy, Central Processing Unit (CPU), and buffer space [115]. To avoid resource consumption and ensure the encounter among devices, there is a need to predict user encounters. The literature for encounter prediction is quite limited. One of the works used the concept of probabilities for encounter prediction. The reliability of probabilities is low, and the outcomes through these techniques are not significant. Hence, there is a need for some machine learning techniques for encounter prediction.

This work has contributed to the research by presenting a method for encounter prediction among different users carrying mobile phones. A supervised machine learning technique is used for encounter prediction that uses a random forest model. The work predicts encounters among different users having mobile devices acting as nodes. The novelty of the work lies in the reduced resource consumption that it offers. In the previous methods [120] , the resources are consumed for continuous monitoring and calculating if a nearby device can be found for data transfer. In this work, with the future encounter prediction, there will be no need to keep scanning for other users endlessly, thus saving on power and computational resources, i.e. mobile battery, and CPU, etc.

The effectiveness of the proposed approach has been tested by deploying an experimental setup to obtain encounter prediction. Furthermore, the obtained results have been compared with a benchmark model and improved upon the former.

An optimized random forest model for predicting encounters among mobile nodes has been used in the proposed approach. The proposed work determines the exact location, and time, at which the encounter will occur with a specific user. The objective of the research was to reduce the required resources for effective communication.

4.2 System Model Encounter Prediction

The mobility traces of a user are used to track his future location, and then encounter information is being estimated. The proposed work uses an optimized Random forest model for encounter prediction. The random forest model is based on the decision trees that store the mobility history of each user. Whenever a user visits a new location or an already visited location at a different time cycle, the tree is updated to reflect the new pattern. In this manner, the future locations of different users, stored in the sub-trees as Spatio-temporal associations, help predict whether users can encounter a specific location and time. Figure 4.1 shows an abstract representation of the proposed encounter prediction model.

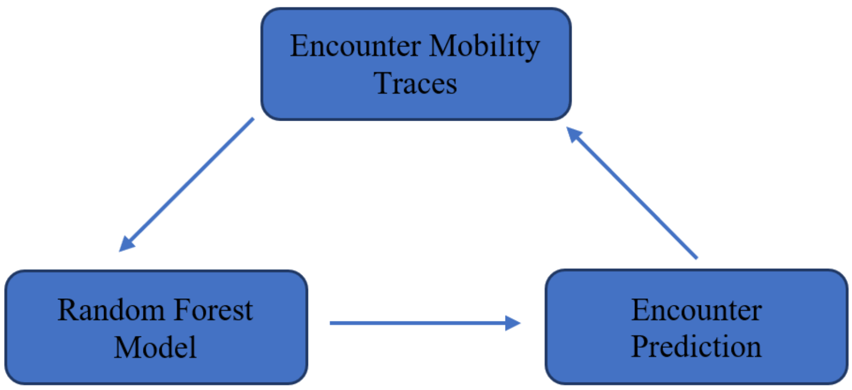


Figure 4.1 Overall working of the proposed model

After calculating the mobility traces of nodes/users of the network, the encounter prediction is carried out using a random forest model. Nodes, in this case, are the mobile devices being carried around by people. The prediction step is followed by a comparison of the estimated encounters with the actual encounters.

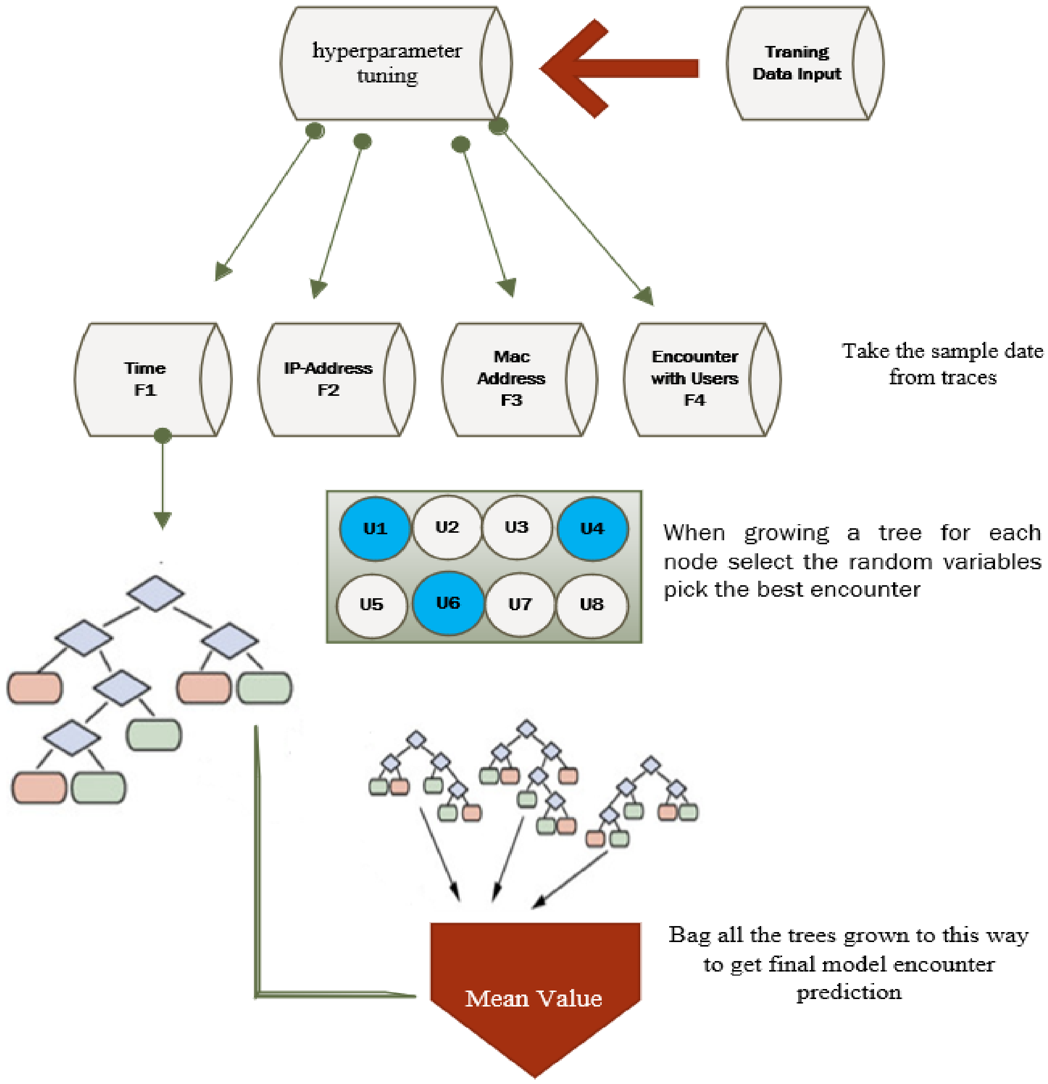


Figure 4.2 Flow diagram of Random Forest encounter prediction model.

Upon a user's visit to a new location, the location is added to the mobility preference of the user. The mobility history update is carried out by adding the location and the time of the visit to the decision tree. If the time and location of a visit match with an existing occurrence in the mobility history of another user, then an encounter is registered. These steps are also shown in Figure 4.2 Flow diagram Random Forest encounter Prediction model

4.2.1 Random Forest Model

The random forest model is based on the concepts of decision trees, where each tree stores information and grows by a specific parameter. It is used for classification and regression

[72] and is based on bagging and random methods. Bagging refers to the idea of constructing a group of learners, each trained on a bootstrap subset obtained from the database. Bagging refers to the independent growth of the trees. For making the predictions, the data from different ensembles of the decision tree is accumulated, and the final prediction is made through simple majority voting among trees of the forest. Random forest trees produce quick output, are easy to implement, and can deal with numerous input variables without the problem of overfitting. In this model, each tree in the forest is created when at first, for each node, input features are selected and then further divided. The best split is calculated based on the input features in the training dataset. The tree grows further in the same manner without pruning. The overall outcome is calculated by taking the majority among the individual outcomes of the trees.

The construction of learner trees in the random forest model can be understood through the following example.

Let D be a dataset consisting of N instances. A D_{sn} will be created by randomly selecting k examples from D with replacement. In case N is large, and the condition $k = N$ holds, then it is concluded that D_{sn} holds probably two-thirds instances from D. For making the predictions, a new cluster or group is created from the separate decisions through classification or regression.

Decision trees play the major role of base learners in the random forest model. Decision trees possess the decision-making ability based on the results of a specific condition. After deciding how many trees a forest will contain, each tree is grown up on a distinct bootstrap instance obtained from the training dataset. Each node of the tree represents a condition/decision against the variable that is selected as the predictor. The purpose of establishing the variable is to reduce the residual sum of squares of both the branches (left and right). The nodes at the end possess only a fixed maximum number of examples from which the target value is obtained through averaging/majority voting. In order to avoid having correlated trees in a forest, the process of choosing the best splitting predictor in each node of a tree is altered. This variable plays a role in determining between the only m randomly selected predictors, i.e. select a random subset of the original n-dimensional problem.

4.2.2 Hyperparameter Tuning

The number of features is denoted by the hyperparameter *F,* which is used in the process of tree induction and selected randomly at each node. The interval [1 ... N] is selected for the hyperparameter *F* where *N* shows the feature space dimensionality. A smaller value of *F* reflects strong mean randomization, which is possible by selecting the accurate number because randomization strength is controlled through the *N* number. A tree structure split (such as splitting criteria is used for feature) randomly choose a feature among all available feature in a case if $F = 1$. On the other hand, if the value of $F = N$, then split selection does not introduce any randomization, and the tree induction process is used to grow each tree. By using the bagging process, randomization is introduced in this specific case. The main contribution of this research is to study the accuracy of the Random Forest concerning hyperparameter *K* on different problems of machine learning.

The main objective of the experiment is to know Random Forest accuracy evolution concerning the different *k* values. In order to know the effect of hyperparameter tuning on the random forest, the role of various values of the *P* must be known. In the random forest encounter experiment, all the random forests are induced by setting the number of trees to 100. The reason behind choosing this value is that it is already given in Jupiter Notebook, which showed that it is the most reasonable value to grow a perfect random forest. Firstly, each dataset will be divided randomly between training and splitting datasets while testing sets are already available. The training and testing data will be two-thirds and one-third of the original dataset, respectively. The split will be shown as a training and testing dataset where $i \in [1 \ldots 50]$. The Random Forest algorithm has been run for each training dataset as *F* in $[1 \ldots f_n]$, where f_n denotes the total number of features. Nonetheless, basically for computational reasons, the definition area of *F* has been tested for a portion of the dataset for which the size of the component space is too huge, so that estimations of *F* have been picked at customary stretches among feature number ‘1’and feature f_n.

Important Hyperparameter: The main reason to use the hyperparameters in the random forest is to enhance the predictive strength and fasten the model. Let’s look at the built-in function sklearns’s hyperparameters of the random forest.

Firstly, the number of trees is denoted by the *n_estimators* hyperparameters. The performance is increased, and prediction becomes more stable through a higher number of trees, but the computation becomes slow. In the experiment, the *n_estimator* is set to 100 according to the results.

The *max_features* is another important hyperparameter that shown the maximum number of features it uses for the normal distribution of the random forest model.

Another important hyperparameter is *min_sample_leaf*, which is used to split an internal node based on the required minimum leaf number.

To determine the allowed number of processors, the *n_jobs* hyperparameter is used. If this parameter contains 1 then it allows only one processor. If the *n_job* parameter value is '1', then unlimited processors are allowed.

The output of the model can be replicable through the *random_state* hyperparameter. The *random_state* definite value is always producing the same result of the model if the same training set and hyperparameter are given.

At last, the Out of Beg (OOB) Score is a cross-validation method in the random forest which is also called OOB sampling. Again, one-third of data is used to evaluate the performance which is not used to train data in this sampling.

4.2.3 Random Forest Encounter Prediction Model

Before formalizing the model, some definitions are being presented. First, select some suitable parameters for the dataset, and these parameters help us improve the accuracy of the random forest model. Throughout the document, suppose that the given training sample data input features $F_n =$ of the dataset with $[0,1]^e \times$ R-valued random variables of encounter ($e \geq 2$) with the same distribution as an independent generic pair F_n. For fixed $u \in [0,1]^e$, the goal is to estimate the classification function $r(u) = E[F|u]$ using the feature F_n.

***Tree Generation**: -*Formally, a random forest is a predictor consisting of a collection of randomized base decision trees [121m ≥ 1][, where E_1, E_2,... are I,i.d. outputs of a randomizing variable E. These random trees are combined to form the aggregated classification estimate.

$$r_n(u, F_n) = E_E[u, E, F_n] \tag{4.1}$$

where E_E denotes the expectation of an encounter concerning the random parameter, conditionally on users u and the features of dataset F_n. In the following, to lighten notation a little, the dependency of the estimates in the sample will be omitted, and write for example $r_n(U)$ instead of $r_n(u,F_n)$.

The random forest model for encounter prediction consists of a set of random decision trees. Each j^{th} random tree r_j has a co-ordinate U and variable E which determine the location for performing cuts for the new tree. In equation (4.2), n is the size of the sample and $\mu_n(U,E)$ is a time whose occurrence resulted in the cuts for the new tree. Let $C_n\,(U, E)$ is the cell obtained after each random partition. Each random tree will be:

$$r_j = \frac{\sum_{j=1}^{n} O_j 1[u_j \in C_n(U,E)]}{\sum_{j=1}^{n} 1[u_j \in C_n(U,E)]} 1_{\mu_n(U,E)} \tag{4.2}$$

where O_j is the output of the tree.

Assume the data is spread in a Euclidean space, and each random tree explores a sub-section of that space. Each tree caters to the region of the search space in a discrete random manner. The selection of location sub-space is itself a random process. The selection of random, sub-space is justified by the notion that is combined with the effect of bagging will result in improved performance. In equation (4.3), the performance of the random tree depends on the selection of sub-space and the cut of the tree applied over successive steps. The time $\mu_n(U, E)$ is:

$$\mu_n(U,E) = [\textstyle\sum_{j=1}^{n} 1_{[u_j \in C_n\,(U,E)]}] \tag{4.3}$$

In equation (4.4), after considering the final expectation E_f, the estimate for each random tree will be:

$$r_j(u) = E_f[r_j(U,E)]\left[\frac{\sum_{j=1}^{n} O_j 1[u_j \in C_n(U,E)]}{\sum_{j=1}^{n} 1[u_j \in C_n(U,E)]} 1_{\mu_n(U,E)}\right] \quad (4.4)$$

In the next step, the correlation among different trees is calculated. The correlation of different trees is calculated using equation:

$$W = \sum_{i=1}^{P} R_i\,(U + \in)/P \quad (4.5)$$

where W is the variable of interest that is to be predicted, p is the no. of estimators used in the search process, R_i is the classification rule set for predicting the value of W. Ui belongs to set U, a learning set. $\in$ is a zero-mean random noise in the data. The definition $\in$ is given in equation (4.6).

$$\in = \beta[\in|U] = 0 \quad and \quad \beta[\in^2|U]\ is\ infinite \quad (4.6)$$

For comparing the results obtained through Gaussian and Random Forest model, the accuracy of predicted encounters is calculated at different locations through both approaches. The accuracy assessments of predictions made by both models were done using the Mean Absolute Error (MAE) of encounter prediction. The accuracy was obtained by comparing observed encounters with the predicted ones. In equation (4.7), MAE obtained predictive models' accuracy by calculating how close each prediction ratio was arrayed around the original data.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (4.7)$$

where x_i is the encounter prediction made for each location, and y_i is the target encounter value and n is the total number of locations for which encounters are predicted?

4.2.4 Gaussian Model

The Gaussian model is used as a benchmark to compare with the proposed random forest model to have a more detailed validation of results. The author has used the Gaussian model [112] for encounter prediction and model diagram of Gaussian process shown in Figure 4.3. The figure shows a Gaussian process model is a collection of random variables. Any finite set consisting of these variables possesses a joint Gaussian distribution. This model can also be defined as a distribution over functions. These functions are the mean and covariance functions.

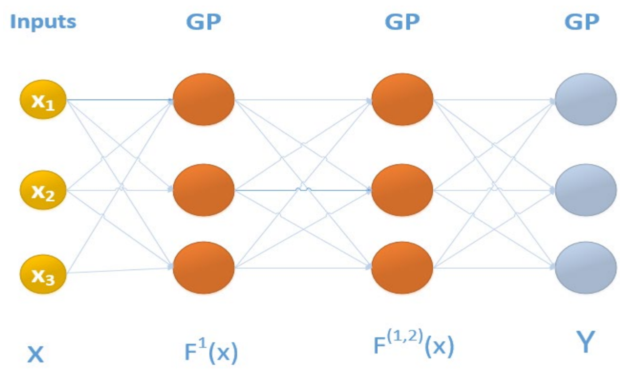


Figure 4.3 Model diagram of Gaussian process

The classification made by the Gaussian process is non-parametric. This method is mostly used for Opportunistic Mobile Networks (OMNs), where connections are sparse.

A Gaussian process is represented through a covariance or mean function (or the kernel) which is:

$$f(x) \sim GP[m(x), k(x, x')] \tag{4.8}$$

where $m(x)$ is the mean function $k\,(x, x')$ is the covariance function for a real process $f(x)$.

For training, the Gaussian process model require a matrix $X = [x_1, x_2, x_3, \ldots\ldots, x_n]$ the size of matrix X is usually $n \times d$ and represents the data points. The model also requires a vector Y, that is: $[y_1, y_2, y_3, \ldots\ldots, y_n]^T$ for training dataset, having size $n \times 1$. f is usually the latent function, and $f(x_i)$ is the predictive class membership probability. The Gaussian model performs mapping of latent space to the observation space, which is usually non-linear.

The kernel function is used to change non-linear input space into space which is represented by various dimensions in a manner where the output of the problem can be shown linearly.

The major problem of this model lies when a probability for a given test point X^* is to be predicted. Where X^* is a member of one of the mentioned classes. The positive class membership probability $p(y = C_+ \mid x)$ is represented by using the following equation:

$$p(y = C_+|x) = sig[f(x)] \tag{4.9}$$

Where f is a latent function $f: R \to R$ which is mapped into the interval [0, 1] through a sigmoid function such that $sig : R \to [0, 1]$, hence, *sig* is a sigmoid transformation function x is a data point belonging to the finite dataset.

The classification through classifier is made using supervised learning. Hence, every data point also contains a class label: $y_i \in \{C_+, C_-\}$

For calculating the predictive class membership probability, at first, the marginal probability is calculated by using the formula given by:

$$P(Y_*, X_*, Y, X, \Phi) = \int P(Y_*|f_*)\, P(f_*|X_*, Y, X, \Phi) df_* \tag{4.10}$$

where, Φ represents the hyperparameter on which the covariance function depends.
In the last, the predictive class membership probability is calculated as:

$$P^* = P(y^* = C_+|X_*, Y, X, \Phi) \tag{4.11}$$

4.3 Algorithm Explanation of Random Forest and Case Study

In this sub-section, the working of the model has been explained through an example. The selection of sample data representing a sub-space is being carried out through bootstrapping. A step-by-step explanation of the algorithm and a working example is presented and before the algorithm is explained in the following steps. The symbols which are being used in the model are explained in Table 4.1.

Table 4.1 The symbols which used in the model

Symbols	Meaning
D	Data (search/data space)
D_t	Training data partition
D_{te}	Test data partition
D_{sn}	Selected sub-space out of the training data
n_{sn}	No of samples in the selected sub-space of the data
O_j	The output of the tree
U_j	Is the co-ordinate of each node
$C_n\,(U,E)$	Cuts of the new tree, in short, denoted as D_{sn}
U,E	The co-ordinates of r_j
r_j	The set of the random classification output
$\mu_n(U,E)$	The time of each co-ordinate
J	Varies $1 \leq j \leq 3$
E_f	Is the estimation function which estimates each random tree $r_j(U,E)$
Mi	μ_i belong to set of μ, which are learning set
R_i	Ri is the rule been set for predicting the value of W

4.3.1 Training Stage

The training/construction of each tree is carried out by first selecting the random subset of samples from the training Data (D).

The training of the random forest model is carried out for each tree independently by selecting samples from the data at random based on the E_i , i.d. distribution, as defined in the case study. The features for training from the set of F_n are also selected in the same manner. The training model selects samples and features using replacement, i.e. some of the samples or features may repeat for the training of each tree of the forest.

At the training stage, the tree building is carried out by following the procedure:

1 By counting samples in the randomly chosen sub-space $C_n(U,E)$, let's call it n_{sn}.

2 The concept behind decision tree training is the reduction in entropy *S,* and maximization of information gain.

3 For each selected feature(s), compute the information gain of the features and compute the difference. For example, for the encounter column/feature, compute the entropy and information gain.

4 The coefficient of variations is used to put a threshold on further branching or if a node has less than a specific number of samples, then stop further partitioning and declare the node as a leaf node.

Where, mse is the error of classification for the node, samples show the number of data points/samples being decided by the node, and value is the average value of the data being represented by the node. In this simulation, the minimum no of samples in a node was set to 1. Hence it leaves only contain 1 sample.

4.3.2 Testing Stage

The model checks all the trees by using the estimation function, which will estimate each random tree, and the majority value of the output of all the trees is output as the predicted value.

Equation (4.2) is used to check the accuracy of each tree, here compute the ratio of correct encounter output to the total number of encounters present in the data, the equation only takes into account the encounter that happened at a specific point in time, where time is represented by $\mu_n(U,E)$.

The aggregate estimate of the prediction is obtained by factoring in the prediction of all the trees, equation (4.3) is being used for this purpose. In this step, simply a majority of the predictions made using equation (4.2) is being computed.

4.3.3 Validation of Output

In the first step, equation (4.2) is being used to evaluate the location feature of the number of users present in the randomly selected sampled data C_n (data partition made through bootstrapping). Here, the location has been taken as the first feature, but it is important to mention here that as the features are selected in the random sequence, the day of the week or time may be selected as the first feature in some of the trees. This makes different trees of the forest uncorrelated, which is one of the hallmark characteristics of random forests. The u_j is one such selected feature among the set. The denominator is the total no. of users at the location irrespective of the time. The notion of time gives rise to the encounter information. The $\mu_n(U,E)$ is the time information about the presence of the user at the selected location. The O_j is the output of the tree.

4.4 Dataset

The University of Southern California (USC) dataset traces are used to evaluate the random forest model for encounter prediction. These traces were real-time mobility measurements. Here, the campus environments are used since they are comprehensive with many active users and associated with sufficient location samples. The data traces are from six different buildings on the campus. The dataset consisted of users' profiles, i.e. their id, information about WiFi associations, i.e. IP address, MAC address of the device, time, and date for a particular connection. The mobility traces were of different months and were recorded for different week, days, including the weekdays and weekends. After the pre-processing of data and noise removal, for evaluation, the traces for January are selected. The mobility traces in the dataset were divided based on the locations. The resultant traces from January were divided into two parts. The first part of the dataset was trained using the Gaussian and random forest. For validating the Gaussian and random forest model, the second part of the dataset is used. Figure 4.4 represents an instance of the dataset used for the evaluation of both models.

	Time	IP_Address	Some_value	MAC_Address	Month
1226	2004 Jan 1 00:03:22	1	5	xx:xx:xx:62:1d:60	Jan
1227	2004 Jan 1 00:03:57	1	4	xx:xx:xx:62:1d:60	Jan
1228	2004 Jan 1 00:07:48	1	5	xx:xx:xx:62:1d:60	Jan
1229	2004 Jan 1 00:10:03	1	4	xx:xx:xx:62:1d:609	Jan
1230	2004 Jan 1 00:20:39	1	5	xx:xx:xx:62:1d:60	Jan
...	...	...	...	...	...
726336	2004 Jan 31 23:51:25	2	5	xx:xx:xx:65:c:21	Jan
726337	2004 Jan 31 23:53:35	2	5	xx:xx:xx:26:c1:48	Jan
726338	2004 Jan 31 23:55:05	2	3	xx:xx:xx:65:c:21	Jan
726339	2004 Jan 31 23:56:31	2	5	xx:xx:xx:19:4e:24	Jan
726340	2004 Jan 31 23:58:50	3	15	xx:xx:xx:18:79:e3	Jan

Figure 4.4 Instance of the dataset used for evaluation

4.4.1 Encounter Prediction Case Study

This section presents a case study to show the application of a designed random forest model over the user's mobility dataset, i.e., the USC dataset [37].

4.4.2 Cleaning and conversion to ordinal data

The data is arranged in the columns of S. No. of the mobility transaction, date and time, IP address as a location marker, MAC address to identify a user. After reading the data, it is first cleaned and converted to ordinal data to make it suitable to be used in the proposed random forest model. A piece of the data is shown in Figure 4.4. The data is cleaned and converted to a sorted temporal form. As a sample few records are shown in Table 4.2.

Table 4.2 Sample data of user 1

Day of the Week	Time (Hour no.)	Location
1	7	3
1	8	3
1	9	2
1	10	2
1	18	3
2	8	2
2	9	2
2	18	1
2	19	3

A user's sample movement/presence activity is explained here. Suppose a user is at home (location 3) at 07 hours and stays there till 08 hrs. Then leaves for work between 08 and 09 hours and reaches the workplace (location 2) at 09. He/she stays in the office until 17 hours and returns home (location 3) at 18 hr. In the same manner, activity for the rest of the day and all week and month is being recorded for each user. A sample data of another user is shown in Table 4.3. Here, user's have been assigned the labels user 1', 'user 2' based on the unique MAC addresses of their devices.

Table 4.3 Sample data of user 2

Day of the Week	Time (Hour no.)	Location
1	7	5
1	8	5
1	9	2
1	18	2
2	8	2
2	9	2
2	18	4
2	19	6

Here, the day of the week represents day no., and time represent hour no., in the example, the time granularity has been set to 1 hour, i.e. time has been rounded to the nearest hour, while in the simulation of the results, it was set to a more practical level, at 1 minute. The time has been rounded to the nearest hour, For example, the events between 07:30-08:29 are mapped to 08 hours, meaning 08:00, and the 08:30-09:29 are mapped to 09:00.

The IP addresses which provide location information have been assigned a countable range. It is important to mention here that the IP address gives the network information and a unique ID number of each device, Here the location represents the network information. For example, If the two users are connected to the same network and assigned the IP addresses, 128.125.214.243 and 128.125.214.240. It is trivial from the example, that they both are connected to the same network. In this manner, each network has been assigned a location no. and in the rest of the thesis, the location has been quoted as the same countable number. Sorting the samples as per the time is carried out to map user movements for temporal prediction of location.

4.4.3 Mapping of Temporal Data for Classification

The data in the sample tables above represents the time-space relationship of the user's presence. As the data will be used for the prediction of his/her next location in the time axis, the temporal data is required to convert to classification data. The user's next location is to be predicted based on his current time and location. In this step, the temporal data is mapped in such a manner that given the event at time t, its prediction will be t+1. For the random forest model to make predictions in time, the data is prepared accordingly and use for training. The conversion of sample data of user 1 is shown in the Table 4.4.

Table 4.4 Data user 1

Day of the Week	Time (Hour no.)	Location	Next Location (y)
1	7	3	3
1	8	3	2
1	9	2	2
1	10	2	2
1	18	3	3
2	8	2	2
2	9	2	2
2	18	1	3
2	19	3	3

Here, the column next location serves as the output/target/class label. The data samples from Table 4.4 are taken for the training of the RF model.

4.4.4 Training of Random Forest model with user's mobility data

In the data presented in the previous sub-section, the last column (next location) of Table 4.4 is set as the target (class label) in the classification training phase. The next location is required to set at the target (y) during the training phase to predict a user's next location based on their current location and time.

A separate random forest model will be constructed and used in the prediction process for each user profiling. For training of the model, the selection of data samples and features is done randomly made, as explained in the previous section. A sample trained tree from the forest is provided in Figure 4.5.

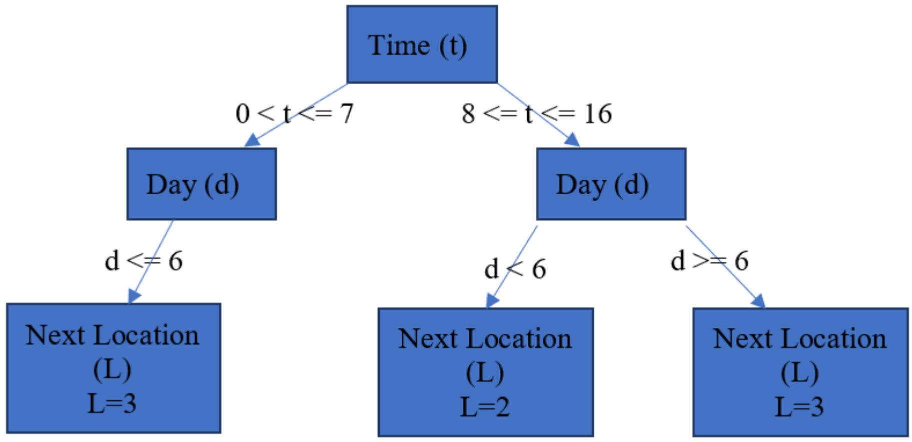


Figure 4.5 A sample trained tree

4.4.5 Interpretation of the Model

The model is built for the prediction of the next location of the user. In the Figure 4.5 the ranges of each attribute/feature on each branch are chosen as the mid-point where most samples from the data will belong. These thresholds are selected based on the method presented in Random forest model.

For example, if a sample data of (Day 3, Time 9) is given and it's required to predict user 1's next location. Begin from the root node and travel along the branches of the tree. In this case, the time of 09 hours leads us to the right of the root, then the day (day 3 of the week, Wednesday) will take us to the left, and here the next location is given, which has been predicted as 2 (work). Similarly, the user's next location can be predicted for all test samples. The data in the sample tree is only using time and day of the week as features, in some of the trees, location is also picked randomly, and a tree is built accordingly.

After training of model user that data, the user's mobility profiles are obtained, through which the next location of any given user can be predicted. To determine where and with whom a user will encounter, first estimate the next location of the user in question. Then, estimate the future (next hour or any given hour) location for all the users and make a list to users with whom the encounter will take place. This procedure later on, will be presented in the form of a formal algorithm.

The prediction algorithm:

1. Trained users profile $P = (p_1, p_2, p_3, \ldots, n)$
2. Encounter estimation for sample ($u, d, t| U, D, T$), D = [1, 2, 3, …7], T = [68, … 3], U = [1, 2, 3 ... n]
3. For a given value of (d, t), estimate the next location of the given user 'I' predict the next location (l)
4. For all other users, compute the next location, and select the one having a similar next location as in the previous step.
5. These users are possible candidates with whom an encounter might happen.

4.5 Experimental set up for Results

In this section, the proposed work is evaluated for encounter prediction among users. The evaluation is done step by step. At first, the observed results for the encounter were calculated. After this, the encounter prediction results are computed with both approaches. In the end, a comparison among the results obtained through both approaches was performed.

For evaluating the proposed work, the dataset is trained with both random forest and Gaussian processes. At first, the actual encounters were determined from the mobility traces, after which the encounter prediction by both approaches was made. The code for training the dataset was written in python code. The encounter prediction model predicted the location, time, and user id with whom the encounter will occur in this code. The output was in the form:

<user_id, day-of-month, time, location_id>.

4.5.1 Observed encounter of different users of the dataset at specific locations

At first, the actual encounters that occurred among the users are noted in the dataset. These results were calculated only for five users to increase efficiency. In the results, two variables are considered, i.e. day of the week and numbers of encounters that were observed on that day. If an encounter was observed at that hour, it is considered as 1 and as 0 in another case.

Figure 4.6 and Figure 4.7 represent the encounters that occurred in January at location 1 (cafeteria) and location 2 (computer lab). These encounters were observed from the real mobility traces in the dataset. The x-axis shows the (day, encounters occurred) ordered pair of January at which the encounter occurred. Y-axis shows the normalized values of the encounters.

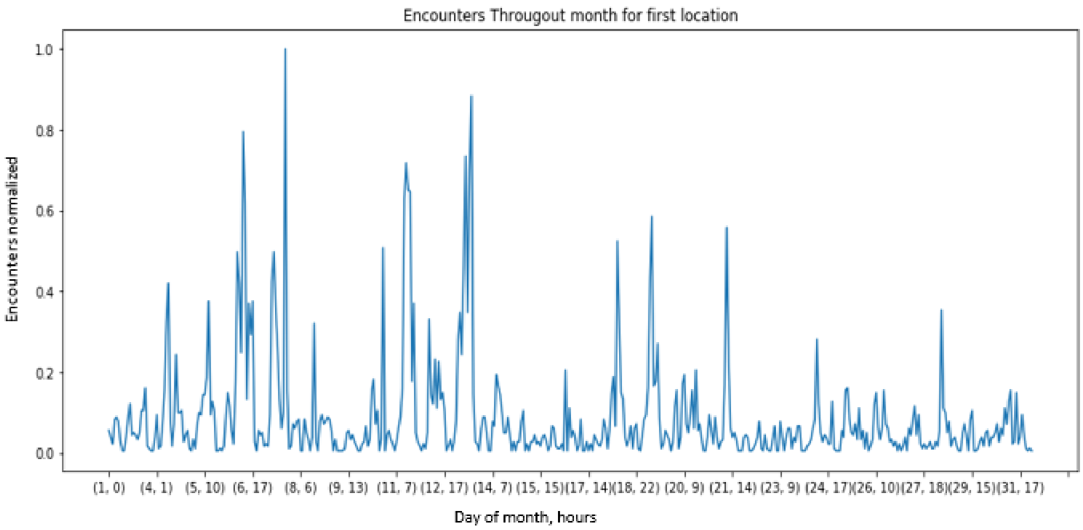


Figure 4.6 Actual Encounters observed in the cafeteria

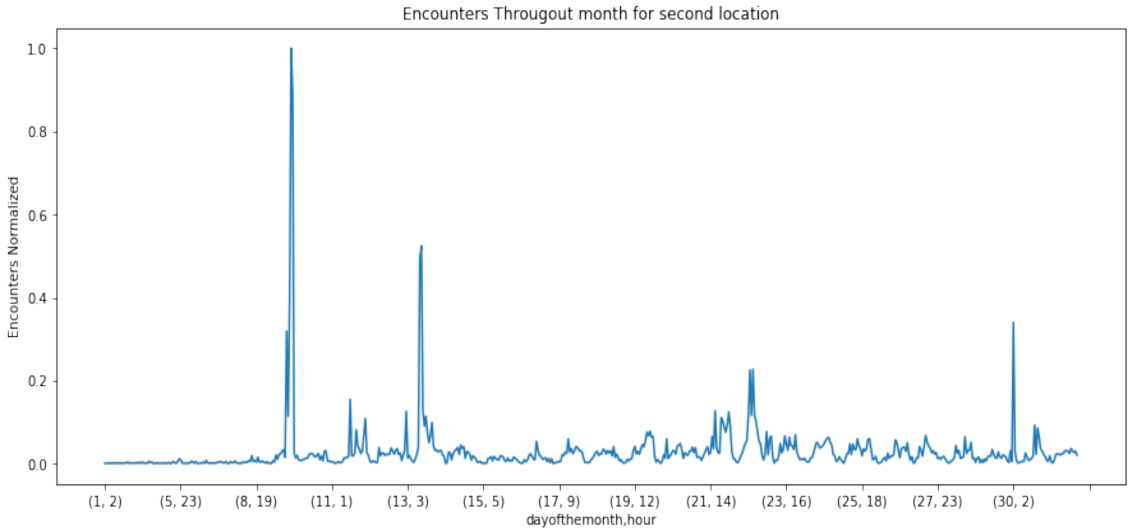


Figure 4.7 Actual Encounters observed in the Computer Lab

The highest normalized values of the encounters were on day 6 and day 24 of January for the cafeteria. As on both these days, 17 encounters occurred. The lowest number of encounters was on day 1, as no encounter occurred on that day in the cafeteria. Variable lengths of the peaks show the different values of encounters that occurred on different days. Similarly, in the computer lab, i.e. in Figure 4.7, the highest number of encounters occurred on days 8, 19. The lowest number of encounters was on day 11, as only one encounter occurred in the computer lab.

4.5.2. Encounter prediction and Hyper Tuning results

This section shows the results of hyper-tuning variables and compares the results of predictions made using the benchmark Gaussian model and the random forest model.

Figure 4.8 and Figure 4.9 both show the encounter prediction vs actual encounter observed through Gaussian Model at the cafeteria and computer lab. The *x*-axis showed the total number of hours in the whole dataset when encounters occurred, *y*-axis shows the prediction values of encounters.

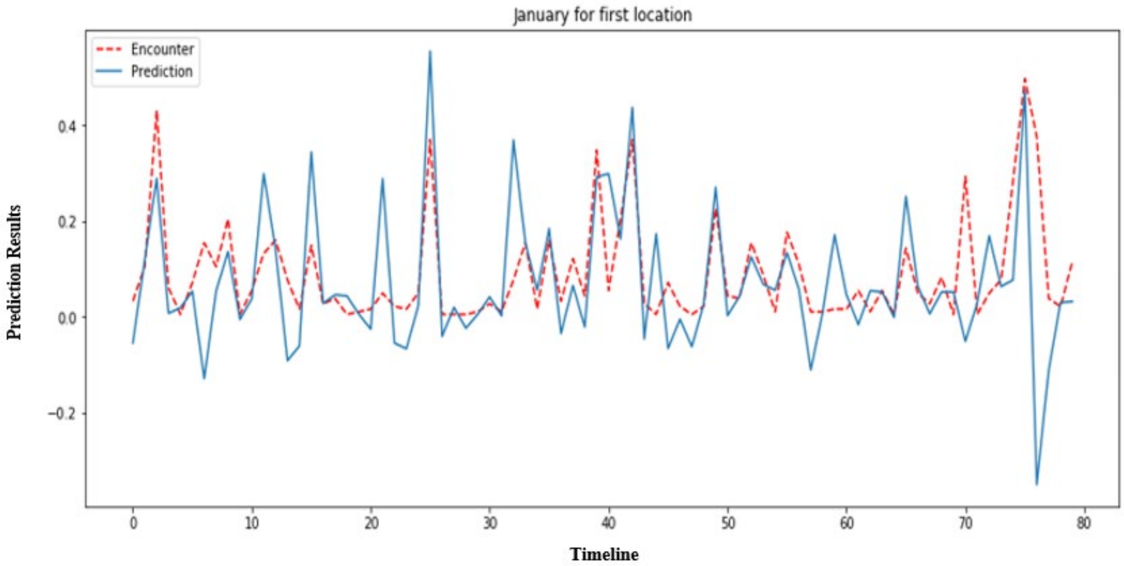


Figure 4.8 Comparison of observed and predicted results by Gaussian model for the location cafeteria

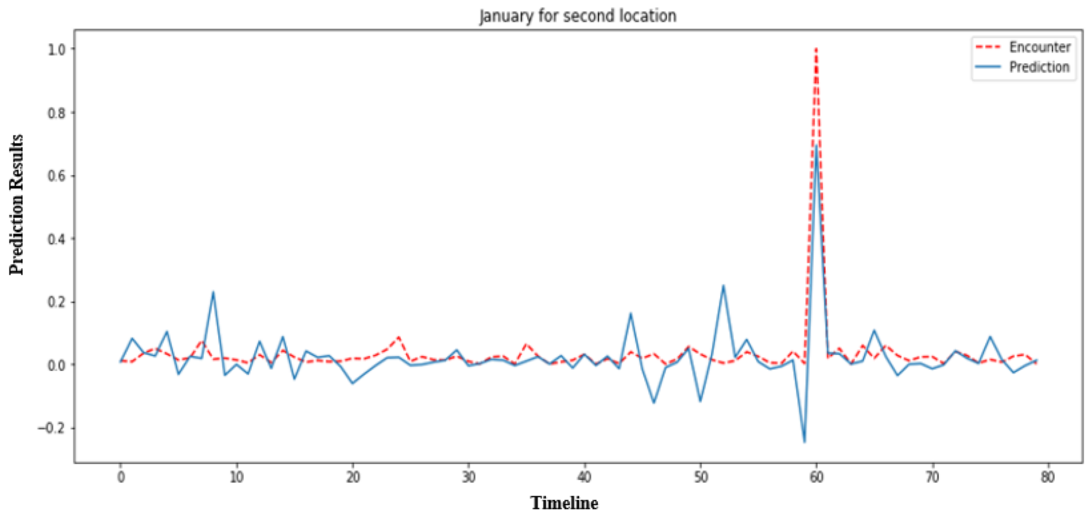


Figure 4.9 Comparison of observed and predicted results by Gaussian model for the location computer lab

Figure 4.10 and Figure 4.11 show the comparison of observed and predicted encounters through random forest model locations: cafeteria and computer lab, respectively.

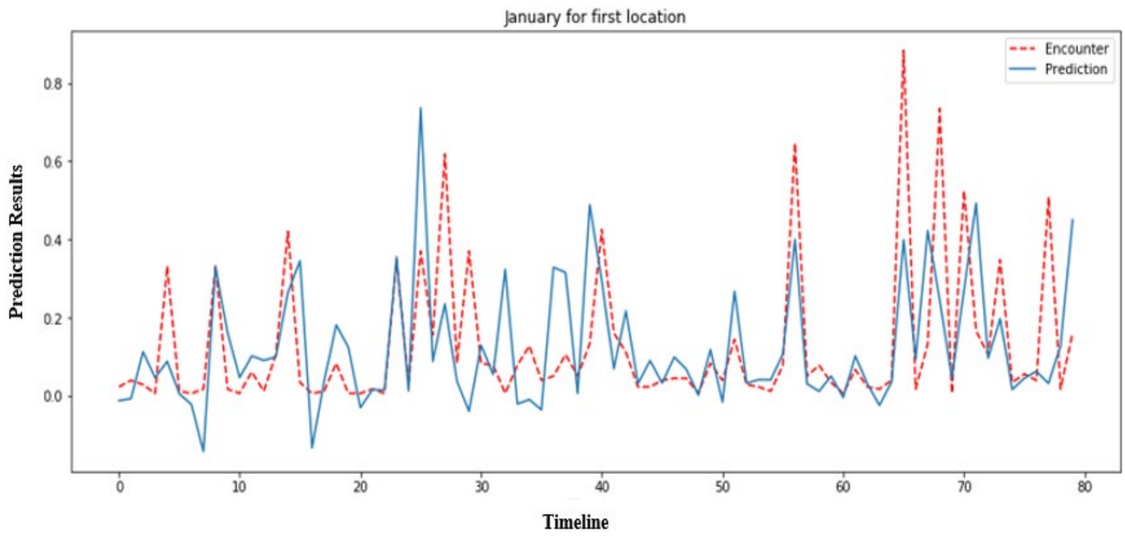


Figure 4.10 Comparison of observed and predicted results by Random forest model for the location cafeteria

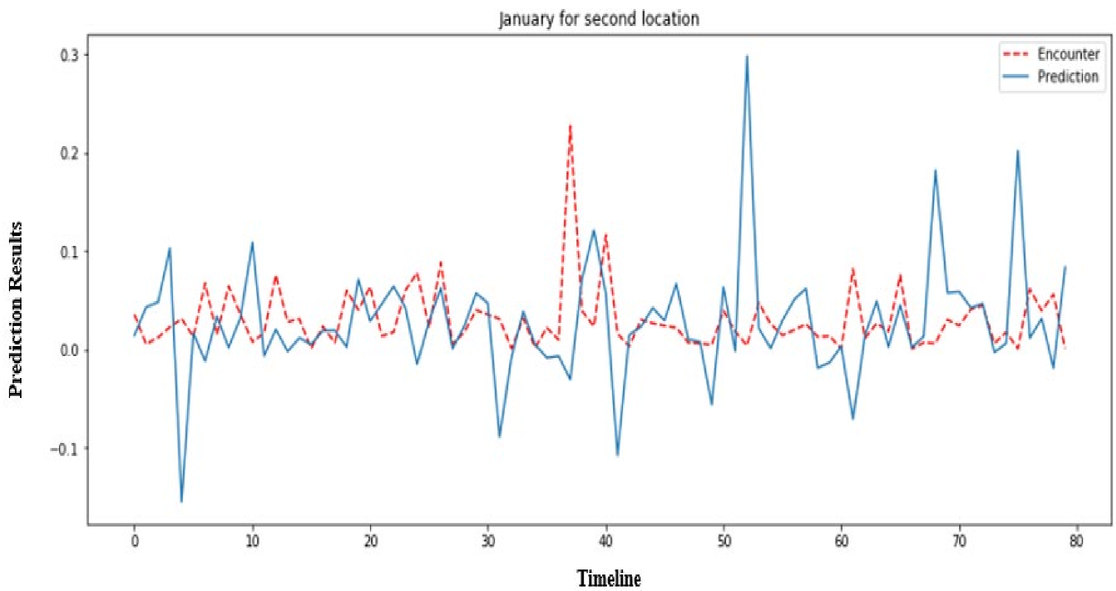


Figure 4.11 Comparison of observed and predicted results by Random forest model for the location computer lab

Figure 4.10 and Figure 4.11 show how close the results for encounters observed and encounters predicted when results were predicted using the random forest model. The maximum total number of hours at which the encounters occurred was 80. These hours are for January. Thus, the predicted and observed encounters for the computer lab were quite close while using the random forest model.

Figure 4.12 shows that when the estimators are increased, the prediction accuracy will be increased, but it depends on the dataset so the random forest model checks the different estimator ranges and select the 100 estimators that are better for the dataset.

To improve the results of the random forest model, I change some parameters based on my dataset, so first, I check the estimators and Max_feature and compare both Max_fearture(auto) max_feature log2. Figure 4.12 shows that the prediction accuracy will increase when the estimators are increased, but it depends on the dataset. Hence, the random forest model checks the different estimator ranges and selects the 100 estimators that are better for the dataset.

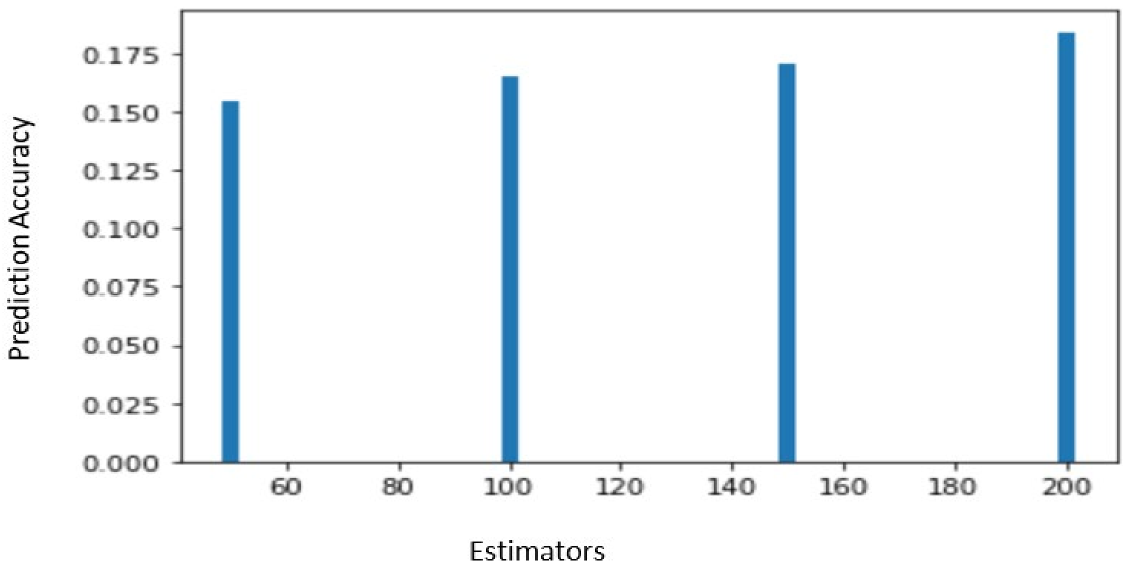


Figure 4.12 The prediction accuracy vs estimators

Figure 4.13 shows Max_fearture(auto) and max_feature log2 used for the normal distribution of data, and the plot shows max-feature log2 is better for normal distribution so the max_feature (log2) is chosen.

Figure 4.13 Comparison of Max_fearture(auto) and max_feature log2

4.5.3 Comparison of results through Gaussian and Random Forest Model

Figure 4.14 and Figure 4.15 show the cross-correlation for predicted encounter results through the Gaussian and random forest models. A cross-correlation plot helps to determine whether a relationship exists among two series or not. The results of correlation between the time shift and the predicted results. As at negative values of the time shift shows, the correlation is decreasing and vice versa.

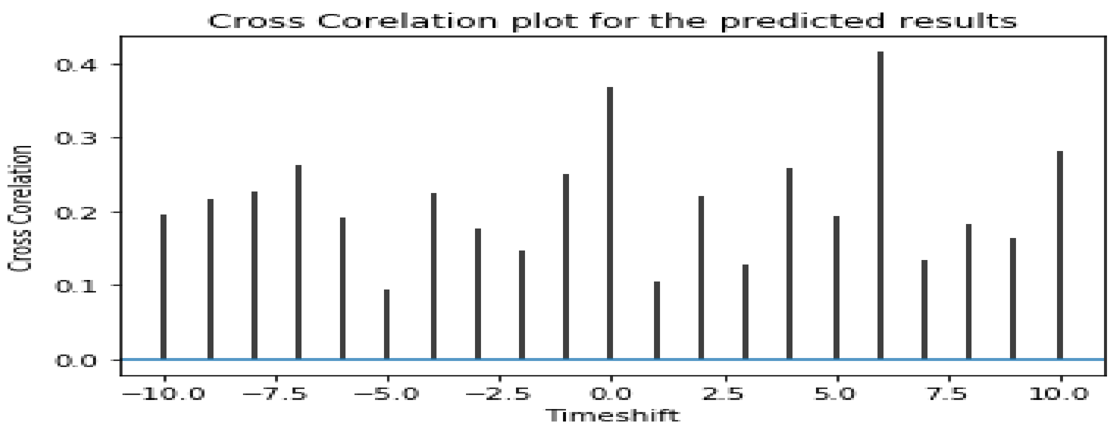


Figure 4.14 Cross-correlation for the results predicted through Gaussian model

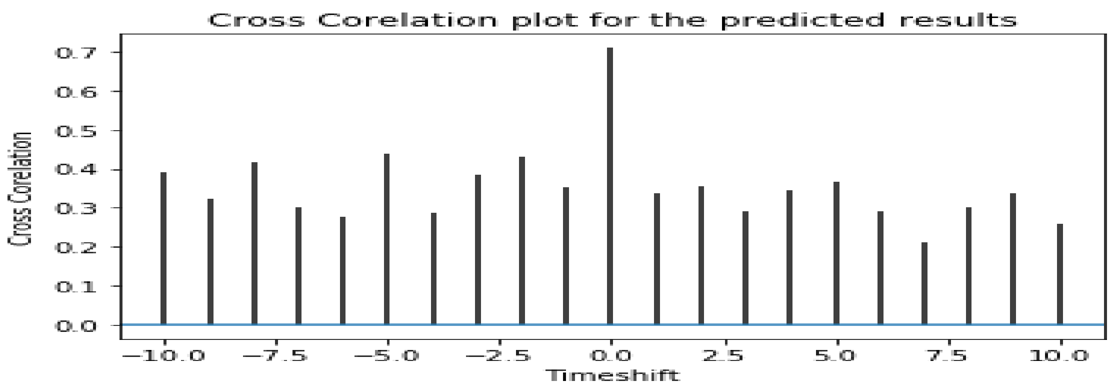


Figure 4.15 Cross-correlation for the results predicted through Random forest model

4.5.4 Accuracy of Gaussian and Random Forest Model

The comparison of accurately predicted encounters at a specific location in different weeks of the month helped determine which model made accurate encounter predictions. Figure 4.16 shows the after-effect of hyper-tuning results of the gaussian and random forest models. Figure 4.16 shows an instance of the output that was achieved after making a comparison among accuracies of both approaches.

Figure 4.16 Graphical representation of comparison among accuracies of Gaussian and Random Forest Model.

The graphical representation of accuracy comparison for results obtained through both approaches (Gaussian and Random Forest) is shown in Figure 4.16. An evident higher accuracy of the Random Forest model can be seen, shown in Figure 4.16. The highest accuracy achieved by the Random forest model was for location 5, which was more than 80%.

Results were obtained through two models i.e. Random forest and the benchmark model (Gaussian Process). Comparison of results clearly showed high accuracy and performance of Random Forest model. The reason for this better performance is because of various factors that differ from Gaussian Process. First, the Gaussian process only deals with the non-linear data problems, while Random forest deals with linear and non-linear data [122] . In the random forest encounter case, a user may have linear and non-linear mobility patterns e.g. in the case of linear patterns, the user may have the same sequence for visiting different locations each day. While for nonlinear patterns, the user may sometimes change the sequence of visiting different locations on different days. That's why a model needs a model that can handle both linear and non-linear kinds of data. Secondly, the superiority of the Random forest model is because of its group or cluster-based learning. As it combines different weak learners to form a strong learner. The decision at a specific point is not dependent on the earlier data. Whereas, Gaussian process is a complete model that relies entirely on all training instances. For encounter prediction, the movement preferences of a specific user may vary on some particular day from his daily routine. Hence, there is a need to make decisions based on the recent mobility patterns done by the random forest model.

4.6 Summary

Human mobility patterns are quite helpful in the data transmission process among different nodes. The mobility patterns help to define user behaviour in the form of movement. These mobility patterns help to predict the future location of a user. The knowledge of the future location of users is important for inferring multiple parameters, i.e. next future encounter. In this chapter, the human mobility patterns are used for an efficient, ensuring, and less resource (energy, bandwidth) consuming data transmission process. The research is presented on encounter prediction, i.e. predicting when two or more users will encounter a specific place and time. The exact encounter prediction can help avoid resource consumption used continuously searching for any available encounter. For encounter prediction, the random forest model is used that stored the history of human mobility traces in tree form. For encounter prediction, the model at first predicted the next future location-based of mobile users based on the data stored in decision trees in the random forest model. Based on the similar mobility preferences of different users, an encounter was predicted at a specific day of the month and hour of the day. The encounter prediction was made based on data stored in the decision trees. A mathematical representation of the random forest model shows how the encounter among different users is predicted using mobility history data of different users. The model showed how the trees in the model grow based on the value of a random variable. The steps for determining specific positions for the cuts in the tree were also explained. The positions for the cuts were important for making new trees. A scenario is given to explain the whole process of how the encounter prediction works. For experimentation, the USC dataset is trained with Benchmark Gaussian and Random forest model. The environment for experiments was developed in Python. At first, the observed encounters were noted down. For results, the encounters were predicted through both, Random forest and Gaussian approaches. For results comparison, a cross-correlation was also plotted through both approaches. The accuracies of predicted encounters through both approaches, their graphical representation, and comparison were also shown in the chapter. The results showed that the accuracy of encounter prediction through the random forest model was quite higher than the Benchmark Gaussian model. Even at some locations, the encounters were predicted even with an accuracy higher than 80%.

The next chapter will show how energy consumption can be reduced through the proposed encounter prediction method will be explored.

Chaper 5
Energy-Efficient Data Dissemination Approach using Multiple-Criteria decision

5.1 Introduction

Smart devices are the digital core of our lifestyle for smooth ways of conducting traditional everyday tasks. These devices have many essential features, including usability, ease of use, and environmental awareness, which induce consumers to consider them personal property instead of a digital tool [123]. In addition to the vast number of applications available for work, entertainment, and social networking, mobile devices provide many methods for connecting with the rest of the world, such as cellular networks, Wi-Fi, and ad-hoc mode [124], and these machines have become omnipresent. Thus, this results in a lot of energy consumption. There are already existing models to reduce energy consumption. These also involve several challenges in the form of the nodes in the networks consuming their resources. This chapter will present the Energy Efficient Data Dissemination Approach (EEDDA) using Multiple-criteria decisions. The suggested approach will save energy in a better way by taking multiple-criteria decisions.

In the previous chapter, the random forest model is used for predicting encounters among different users having cell phones. The proposed work determines the exact location, and time, at which the encounter will occur with a specific node. The purpose of the previous chapter was to reduce the resources consumed for continuously searching any available nearby node with which the encounter may or may not occur. In the previous chapter, the human mobility patterns are used for an efficient, ensuring, and less resource (energy) consuming data transmission process. In chapter 4, encounter prediction, i.e. predicting when two or more users will encounter at a specific place and time. A random forest model is used that stored the history of human mobility traces in tree form for encounter prediction. For encounter prediction, the model first predicted the next future location of the mobile user and stored it in a decision tree in the random forest model. The encounter prediction was made based on data stored in the decision trees. The mathematical representation of the random forest model is also provided. The model showed how the trees in the model grow based on the value of a random variable. The steps for determining specific positions for the cuts in the tree were also explained in the chapter

4, Random Forest Encounter Prediction Model section. How the encounter prediction works is also explained with the help of a scenario.

At first, the introduction and the need for an energy-efficient data dissemination approach using multiple-criteria decisions are mentioned. Then some existing works and techniques for energy-efficient data dissemination are listed. The model used for energy-efficient data dissemination is explained along with the mathematical representation of the model. The experiments, results, and comparison with standard techniques are also discussed in detail at Energy consumption model section.

5.2 Background

In the option of data transmission through the internet, energy usage depends on the energy used by nodes at both ends for transmitting and receiving sides to download and upload the amount of network equipment used, such as hops used before the final destination, in between. The option of data transmission through Device-to-Device (D2D) energy consumption depends upon data volume, sending, and receiving sides. Our goal is to find out which communication way is using less energy consumption. In this chapter, the EEDDA uses multiple-criteria decision energy models. The model shows the multiple-criteria decision, and based on encounter prediction results, the node decides to transfer the data through the internet or D2D. The number of recent studies discussing and implementing D2D communications [124] is broad, ranging from those dealing with the discharge of mobile data traffic to those dealing with the extension of network coverage or sharing of content in community areas.

Mobile and wireless networks are part of our lives. More and more people are becoming part of interactive networks to exchange multiple details. A Mobile Ad-hoc Network (MANET) is a decentralised, network-less infrastructure activity that arbitrarily switches wireless nodes. The nodes communicate with nodes in the transmission range directly and interact with others in multi-hop communication. Such networks are rapidly gaining popularity because of their easy integration. Due to the rapid economic growth influenced by industrialisation and globalisation, energy consumption has been steadily increasing. Industry, transport, and buildings are the three main economic sectors that consume a large amount of energy, having the highest proportion of homes. In mobile ad-hoc wireless networks, energy consumption is a significant issue, as most mobile nodes operate on a limited battery resource. The latest models for calculating energy use in mobile ad-hoc networks have shown that both transmission capacity and processing capability are involved in the various components of energy-related costs.

Energy efficiency is of great importance to the research community in expanding the analysis to wireless networks, in general, to cope with the continued growth of energy-demanding applications in scenarios with limited energy resources [121]. In particular, green radio solutions are being studied for potential wireless systems [125, 126], the base station side implements a discontinuous transmission mode (DTX) to reduce LTE communications use. In the LTE-A standard, the discontinuous reception/transmission (DRX/DTX) mechanism is also defined to allow devices to go to sleep when data from/to the base station does not need to be received or transmitted. For MTCs over cellular infrastructures, this function is certainly of particular interest. Liang et al. [127] investigated that, in particular, the DRX/DTX optimisation to optimise devices' sleeping cycles, ensuring their standard of service, but no one worked on multiple-criteria decision approaches.

Based on the discussions mentioned above, surveys and classifies several energy-efficient mechanisms proposed for MANETs. Generally, they can be described based on when the energy-saving is over. A mobile node uses its battery power not only when it actively sends or receives packets but also for some particular potential contact requests from other nodes when it sits idle listening to the wireless medium [18].

Reliability is less in the existing models, and the outcomes through these techniques are not significant. Hence, there is a need for some energy-efficient data dissemination techniques using Multiple-criteria decision

This work has contributed to the research by presenting a model for Multiple-criteria decision energy consumption among different users having mobile phones. Our proposed model will show the Multiple-criteria decision of node takes the decision based on encounter prediction results from data transfer through the internet or D2D or internet. The University of Southern California (USC) traces are used for this research. The decision engine will help to sustain the Quality of Service (QoS) and help to save energy. The novelty of the work lies in the reduced resource consumption that our proposed model offers. The key idea is to minimise the infrastructure-based transmissions by best utilising the devices' mobility to encounter following among smart devices through the short-range D2D tethered communications using the internet of moving things. Based on the rich opportunities for mobile experiences, this chapter has been verified based on the similarity study of real mobility traces. Exist in human society. Hence, the research is focused on reducing energy consumption and sustain the QoS's.

The chapter proposed the best utilisation of the existing telecommunications infrastructures, infrastructure-less mobile social networking in proximity, to transmit the traffic with delay-tolerant features. A network model is proposed to reflect the energy consumption problem to measure the more efficient energy solution. The results of experiments show the D2D is saving energy consumption with the help of multiple-criteria decisions. For the analysis of this research, an experimental setup is used to obtain the results for the multiple-criteria decision energy model.

The conventional world has become a digital world where it is possible to access almost anything everywhere. The digital society, however, is responsible for high-energy consumption.

Several business sectors have benefited from emerging technology with technological developments over the last few years. The trend is to develop more and more inventions to meet our daily demands. Yet another aspect to consider is their safe and effective use. Advances in technology such as 3G, 4G, and 5G, and so on give us advantages but equally cause of high-energy consumption. The energy conservation theory has been introduced to cope with this situation. One of the main issues is to reduce network infrastructure energy usage by introducing various approaches, such as an energy-efficient approach to data dissemination.

In this chapter, first, the energy challenges will be addressed and then how the EEDDA approach checks the encounter prediction results and stability of encounters and takes multiple-criteria decisions.

5.3. System Model for Energy Efficient Data Dissemination Approach.

The elaboration of the system model for an energy-efficient data dissemination approach is given in the next paragraph.

The EEDDA take the multiple-criteria decision by using the encounter probability results and checking the Delay time of the user, transfer time of data, encounter time of senders and receivers, the data size. Then the EEDDA will decide that the data may be transferred through the internet or D2D.

The current system first uses the USC data traces of a user having a mobile phone (node) taken from the building location App-based to reduce energy consumption. An encounter device is used. It will take prediction of encounter traces and IP address stability, then Analyse both. After that encounter device forwards them to the decision engine. Then with the help of the decision engine, data will go through the internet and sustain the QoS or Data travel D2D and save energy consumption. In the last phase, the Energy Efficient Data Dissemination Approach using Multiple-criteria decision works for reducing energy consumption in the overall network.

In Figure 5.1, the overall working of the energy-efficient data dissemination approach will examine the delay tolerance indicator concerning the user who wants to initiate the communication by sending the data. Next, the proposed multiple-criteria decision approach will check the encounter probability between sender and receiver, then compare the encounter probability of both the users with delay tolerance index to ensure that both the user, i.e., sender and receiver, will encounter a given period. Finally, the approach will check the users' encounter time, i.e., sender and receiver. If the conditions are specified after confirming, the model would decide the mode of communication either by D2D or over the internet.

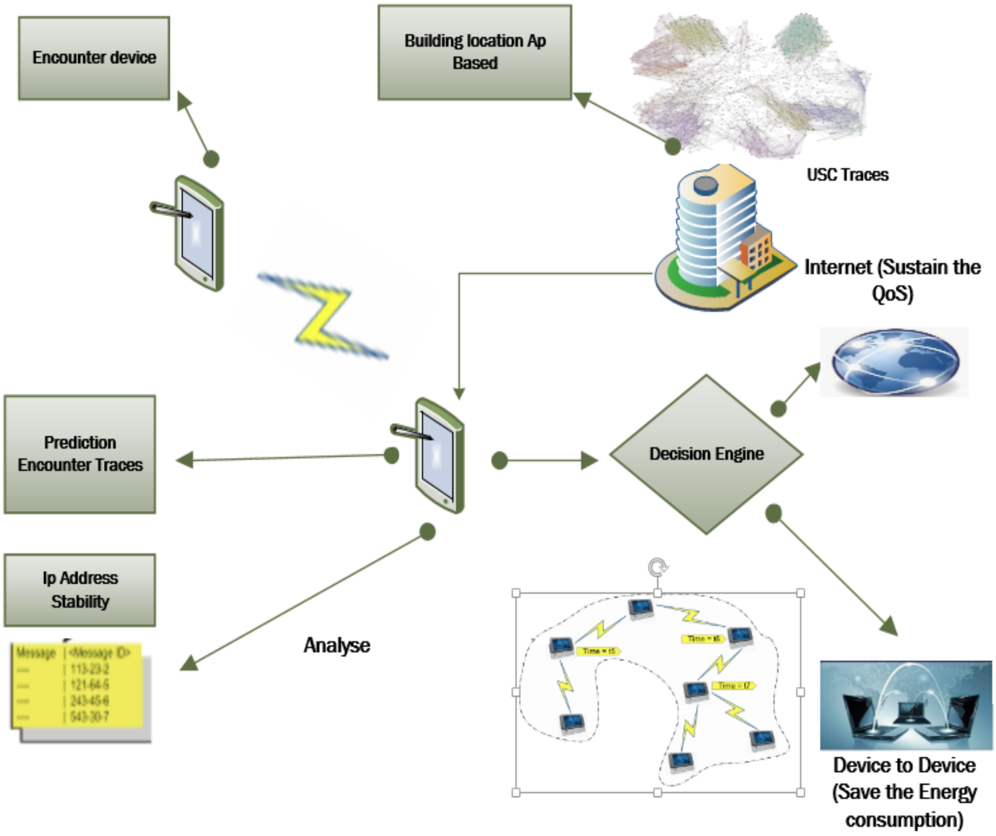


Figure 5.1 The overall working of the Energy-Efficient Data Dissemination Approach.

5.3.1. Energy Consumption Model

To measure the more sustainable energy solution, a network model reflects the energy consumption problem. Let's say that multiple networks are directed by $G = (N, L)$ with N as some nodes and L are a set of edges in the network. In a network, the flow of objects from I to j between n nodes is represented by edges (i, j), $I, j \varepsilon E$. Edges have I_j E a potential and an energy consumption function C_i, j. $S \varepsilon N$ Display the vertex of the sender node, and transmitter $x \in N$. Bandwidth is also based on network link capacity, and they are defined as U_{ij}. Considering the demand and source $Bi < 0$, in between a relay node with demand $B_i = 0$ and a demand destination $Bi > 0$.

For prototyping, assumed a general and simplified energy consumption model for wireless or wired energy dissipation where the transmitter dissipates power to generate radio or line electronics. The power amplifier consumes energy to transmit traffic, and the receiver dissipates energy to receive and process radio or line electronics, as shown in Figure 5.2.

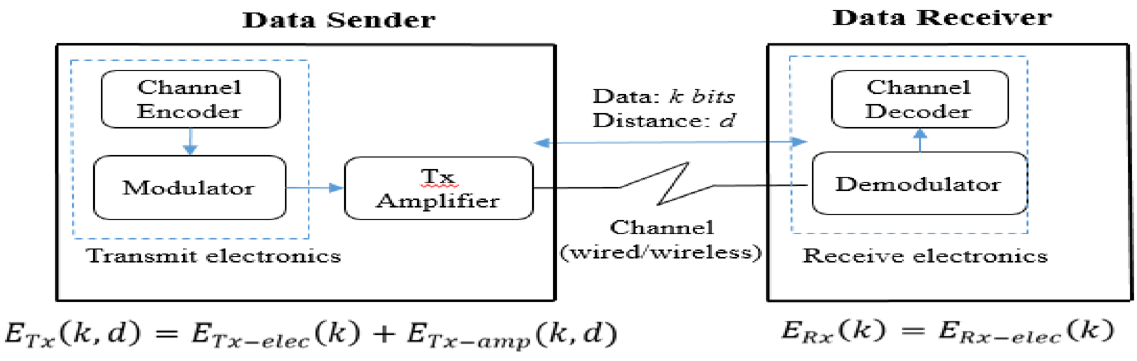


Figure 5.2 A general Energy Consumption Model

Taking radio transmission as an example, by setting the power amplifier accordingly, the power control can correct signal propagation loss. For example if the transmission distance is less than the threshold d 0, the free space prorogation model is used with the attenuation parameter ε_{fx}. Otherwise, the multipath (mp) propagation model is used with the attenuation parameter ε_{mp}:

$$E_{Tx}(k,d) = E_{Tx-elec}(k) + E_{Tx-amp}(k,d) \tag{5.1}$$

For the case of radio transmission:

$$E_{Tx}(k,d) = \begin{cases} k \cdot E_{elec} + k \cdot \varepsilon_{fx} \cdot d^2, d < d_0 \\ k \cdot E_{elec} + k \cdot \varepsilon_{mp} \cdot d^4, d > d_0 \end{cases} \tag{5.2}$$

where E_{elec} is the energy consumed by the transmitter. It depends on variables such as the processing of digital coding, modulation, and filtering signal. As for the power used by the amplifier, the distance to the receiver and the acceptable bit error rate depends on it. It is possible to measure the energy consumption of the data obtained by:

$$E_{Rx}(k) = E_{Rx-elec}(k) = k \cdot E_{elec} \tag{5.3}$$

Based on the above general energy consumption model for communications, two significant and changeable factors will change the overall energy consumption. It is compared to the energy consumed by the electronic components and the signal processing mechanism in the transmitter, receiver, and relay amplifiers, as seen in the amount of data k and the transmission distance. Can the transmission distance d that is being traversed through the networks be reduced? The shorter the transmission distance, in other words, the lower the consumption of electricity. The minimum cost flow issue for the various data volume numbers is clarified, and less energy consumption is measured by moving the data between two locations and nodes. Multiple-criteria decisions based on encounter prediction for EEDDA.

5.3.2 Energy Consumption model for Device to Device and Internet

When sending data through the internet, data will be submitted to the core network and forwarded according to the TCP/IP concept, and, on the other hand, it will be downloaded from the core network. Energy usage depends on the energy used by the transmitting and receiving sides of the nodes for downloading and uploading at both ends. The number of network devices used between routers/hops is used until the final destination. The delay value depends on the bandwidth used and the storage space available to carry data in equation (5.4); I_{Nodes} is the power consumed by nodes for uploading and downloading.

$$I_{Nodes} = max\left[\frac{m}{b_1}, \frac{n}{b_0}(\nabla E_m + \nabla E_n)\right] \tag{5.4}$$

where b_1 and b_0 the bandwidth upload and download of the node that sends and receives. In our case, each node's upload and download bandwidth is 0.1 Mbits/s and 1 Mbits/s for uploading and downloading data to the internet. During data transmission, the difference in power consumption by a node is E_m.

$$C_{bit} = \frac{P_1}{b} \tag{5.5}$$

where n represents as the number of bits being exchanged, b is the highest bandwidth used, and P - 1 is the maximum power extracted by a device. The incremental energy cost of one-bit transfer to model the energy usage of network equipment will also be considered.

$$I_{inc} = n \times C_{bit} \quad (5.6)$$

Also, the overall incremental transmission energy cost should be considered as the amount of node energy consumption and the total node per node. Where k is the number of network devices used during the connexion, in the EEDDA case study, suppose there are three switches, two core routers, and two edge routers in both places. Energy consumption can be estimated from equation (5.7), as each nodes' total energy and energy usage.

$$E_c = I_{nodes} + \sum_{j=1}^{n} I_{inc1} \quad (5.7)$$

The D2D communications technologies have been developed recently, and the smart devices nearby You can directly interface with each other and structure a communication network. Instead of transmission through the devices, data traffic can be offloaded to the D2D network infrastructures such as base stations.

For example, some clients download substances from BSs by approving internet communications, while others might recover substances from their associates. D2D messages, along these lines, significantly reduce the limit requests in BSs and likewise reduce BS energy consumption [125], calculating the D2D energy data sent independently to N users, according to the equation (5.8):

$$DD = R_f * \left(U_p + \nabla_L T_{PX}\right) * (S_{mbps}/A_I) \quad (5.8)$$

U_p = Minimum active power unit 130W, R_f = number of radio frequencies 6, ∇_L = linear transmission dependence factor 4.7, S_{mbps} = Data size 20 Mbps, T_{PX} = Transmission power, A_I = Air interface 4.

5.3.3 Multiple-criteria decision model among traditional and Device to Device network

Algorithm 1 in Figure 5.3 shows the multiple-criteria decision model math. Inputs of the multiple-criteria decision model are some variables which are declared at the start of the model which includes, Users denoted by u_1 to u_n, Time Encounter by user *EP(u)*, the maximum time to transfer denoted by *T*, Delay Tolerance Indicator in days denoted by *DTI(d)* and location of users denoted by *Loc(u)*. The output of our proposed Multiple-criteria decision model is the selection of optimal data transmission mode, i.e. D2D or over the internet between users. The output of the proposed model will be purely based on values of defined variables at the first instance of the model.

Algorithm 1 : Multiple- decision model

Input : Users= $[u_1, u_2, ...u_n]$, t: Encounter Time, EP(u): Encounter Probability, T= Max time to transfer, DTI(d)= Delay Tolerance Indicator in days, loc(u)- Location
Output: Optimal Network Selection

while *true* **do**
    $u_1 \leftarrow DTI(d)$; `// Check DTI for sender user`
    $(u_s, u_r) \leftarrow EP$; `// Check Encounter probability between sender user and receiver user`
    $EP(d) \leftarrow DTI$; `// Encounter is possible within DTI`
    Available$(u_s, u_r) \leftarrow t$; `// Check for enconuter time for sender user and receiver user`
    for $U(u1, u2, ...u_n) \in Loc$ **do**
        if $t \leq T$ **then**
            mode = *Data-Disseminate*(s, d) *using D2D;*
        else
        end
        mode = *Data-Disseminate*(s, d) *using Internet;*
        return *mode;*
    end
end

Figure 5.3 Multiple-criteria decision Model

Operations of the proposed multiple-criteria decision model are explained step by step below:

- First, the model will examine the delay tolerance indicator concerning the user who wants to initiate the communication by sending the data. Line 2 in algorithm 1 illustrates the mathematical form of this statement.
- Next, the proposed multiple-criteria decisions will check the encounter probability between sender and receiver, as per line 3 of the algorithm.
- Line 4 of algorithm 1 illustrates the comparison of encounter probability of both the users with delay tolerance index to ensure that both the user, i.e. sender and receiver, will encounter a given period. The results of encounter probability, its extracted via the Random Forest Model. In the previous chapter, the Random Forest model is discussed and demonstrated.
- Line 5 in the multiple-criteria decision algorithm above will check the encounter time of both the users, i.e. sender and receiver.
- If the conditions specified above met for all the users after confirming in line 6 and 7 of algorithm 3, the model would decide the mode of communication either by D2D or over the internet based on the following conditions.

If encounter time is less than equal to Maximum Time = (data volume/ data transfer rate), the data will transmit via D2D mode; otherwise, data transfer over the internet.

The following examples illustrate the selection of data transmission modes by multiple-criteria decisions in different cases.

For example, the direct data transmission range is 250 Mbps via a Wi-Fi link. In the EEDDA case, it is assumed that the ideal data transmission rate of 200 Mbps by reserving 50 Mbps for obstacles and other communication barriers between source and destination user. Hence:

$$
\begin{aligned}
&D_s = 100MB \\
&D_r = 200Mbps \\
&T_{Th} = 1s \\
&\frac{D_s}{D_r} = \frac{100MB}{200Mbps} = 0.5s
\end{aligned}
$$

As $0.5 < 1$ it means data is possible to transmit within a threshold time. Hence, upon meeting all other conditions, communication between u_s and $u_r$will happen in $\mathrm{D_2D}$ manners.

Now, look at the other side of a pillar, assume the data required to send by $u_s$to u_rWhile communication is 250MB instead of 200MB as per the previous example. Now;

$$
\begin{aligned}
&D_s = 250MB \\
&D_r = 200Mbps \\
&T_{Th} = 1s \\
&\frac{D_s}{D_r} = \frac{250MB}{200Mbps} = 1.25s
\end{aligned}
$$

As 1.25 > 1 so last condition set in the algorithm returns false. Therefore, communication will be occurred over the internet via LTE mode despite all other conditions return true.

5.4 Numerical Studies:

The numerical studies describe the Comparison of Energy Consumption for the Device to Device and the Internet. Multiple-criteria decision model effect on energy consumption based on data size, number of users, delay.

5.4.1 Comparison of Energy Consumption for the Device to Device and the Internet.

The EEDDA Energy-Efficient Data Discrimination Approach is used to minimise energy consumption and sustain QoS. The results provided in section 5.4 prove the minimised energy consumption cost for data transfer through the device to device or internet.

This chapter considers three parameters that affect energy consumption. These parameters are distance among sender-receiver, data size to be sent, the number of users on the network, and the delay of time (the user wants to send the after some time). Distance between two users plays a vital role during communication. The size of the data defines how much data two users wanted to transmit which each other. The number of users also affects the performance due to bandwidth limitations; as the number of users on the networks increases, the delay in transferring the information between two users increases. The importance of the data also defines what the worth of data is. Do you need it on an urgent basis, or can you send it later.

This work is focused on the effects of these parameters on energy consumptions between D2D and through internet communications. Only those parameters are chosen which have a direct impact on energy consumption. The section discusses data size and the number of users on energy consumption between D2D and internet communications.

5.4.2 The Effect of Data Size on Energy Consumption

Data size is another parameter that plays a crucial part in energy consumption in the data transmission process.

The total energy consumption of a specific amount of bit can be calculated using the amount of energy consumed by sending and receiving nodes and all the switching and routing equipment routing data between two nodes. The total energy consumption can be calculated using equation (5.4) which gives us the energy consumptions between two nodes. The parameters m and n define the number of the bits which need to be transmitted between two nodes. The incremental energy cost is given by using equation (5.6), while equation (5.5) provides the maximum energy consumed for a given bandwidth. The total energy for the communication between two nodes on the internet is calculated using equation (5.7). The energy consumption using D2D is calculated through equation (5.8).

To calculate the effect of varying data size on energy consumptions, considering the EEDDA case study where communication is done between two nodes. While assuming the number of nodes in equation (5.7) remains fixed, the effects of direct changing the data size on energy consumption can be seen. The data size is changed from 0.5Tb to 64Tb, and the energy consumption for both the internet and D2D is calculated. Figure 5.4 shows a comparison of energy consumption according to data with varying sizes. The experiment was performed for the same group of people as in the section. The energy is consumed more when data is sent through to the internet. This is because big data size requires more time and energy for uploading on the internet and then for the transmission to the destination. In this, the sender and receiver's data is calculated with various data sizes and calculated the energy with both aspects, transfer through D2D and Internet.

The plot depicts that as the data size increases, the energy consumption in the case of the internet is extremely high compared to D2D communications.

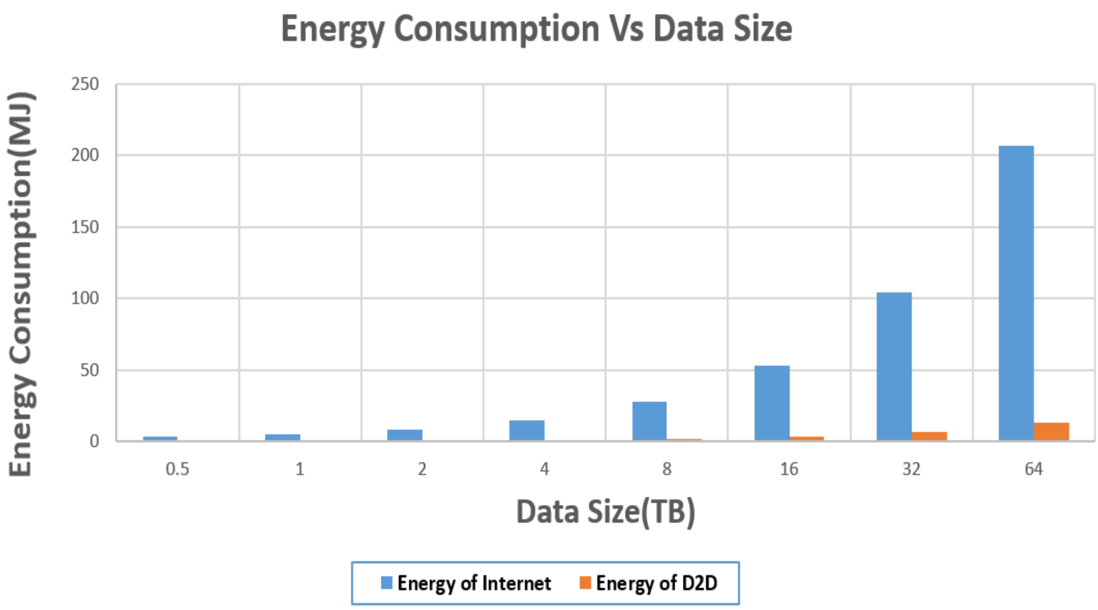


Figure 5.4 The comparative analysis of energy consumption based on Data size

5.4.3 Effect number of users on Energy Consumption

To evaluate the effects of an increasing number of users in terms of energy consumption, utilising the growing number of nodes while keeping the data size fixed in both internet and D2D communications. The effect of the increasing number of nodes is discussed in both cases. The number of nodes is increased from 10 to 200 while considering a campus scenario where users wanted to transfer a varying data size between them. The individual user's energy is calculated and added to calculate the overall power of the users. The number of nodes and other parameters remains fixed for fair evaluations. As the number of users increased, the amount of energy consumption between them grew in an exponential form. The change in energy consumption in the case of D2D communication is lower as compared to the internet. The results of this case study are shown in Figure 5.5. Normally, if the number of users increases, energy consumption transmitting data in the network is high, but still, it is low compared to that consumed for data transfer using the internet.

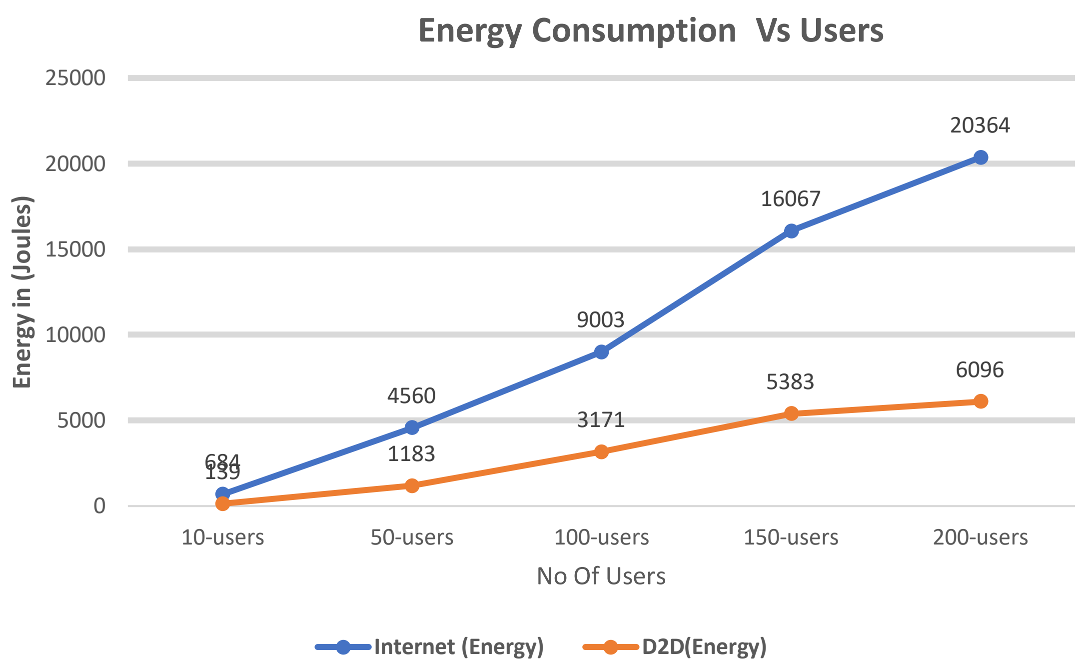


Figure 5.5 The comparative analysis of energy consumption based on No of users.

This plot is based on our energy model. This plot shows that 200 nodes transfer the data through the D2D, then how much energy will use and if the same 200 nodes send the data through the internet, how much energy will be used. These nodes select the data size randomly. Now, it can be seen when ten users send the data through D2D, then 139 Joules J of energy is used in the same scenario when ten nodes send the data through the internet 684 J will use. It clearly shows that the internet uses more than then1/3 energy than D2D; even reach 200 users in D2D use 6096 J energy when the internet uses 203 J energy simultaneously. This shows that D2D is more efficient.

5.4.4 Delay and Multiple-criteria decision model effect on energy consumption

The multiple-criteria decision model is used to select the preferable mode of communication between two nodes. The multiple-criteria decision model considers the traces of the communication to consider to select the mode of communication. The model chooses the mode for a specific time interval between the user, and after that specific time, the model will again decide the mode of selection between two devices. Each data has a different sensitivity. Suppose the time interval in which data must be transmitted is increased. In that case, more users can be physically located in the same location so that the multiple-criteria decision model will select a device. For generating these plots, the USC dataset is used to contain the traces of the user in the form of id, start time, end time, person, and location. The multiple-criteria decision model uses the traces of the users to select the optimal model of communication between using their time and location during a specific time interval. The users are located in 6 possible locations and traces for seven days from Monday to Sunday. The traffic demand between the users is selected using a random size of data between different users.

To calculate the impact of the delay, the EEDDA utilises a multiple-criteria decision model that calculates the optimal way of transferring the data based on the prediction based on a specific time slot and specific location. Each data has a different severity level, and some data needs to transmit immediately while others can be transmitted in several days.

The plot shown in **Error! Reference source not found.** compares 'Energy consumption using multiple-criteria decision model vs delay.' They showed how the delay affects choosing the network to transfer the data and its impact on energy consumption. The multiple-criteria decision models will only allow one permissible connection between two nodes, as given in equation (5.4). Two nodes can transmit data either using D2D or through the internet. The EEDDA model ensures only one connection between two nodes will select an optimal approach to transfer the data through D2D or internet communication. The model will utilise the traces of the users to decide the mode of communication between users. The multiple-criteria decision model will do away with selection for users' requests and also determine the connection activation. A user can be either a D2D user or an internet user at any given time frame. The traceability information is needed for the multiple-criteria decision model for selecting the mode of communication between two users at any given time interval. The decisions upon the possible way of selection are made based on previous traces. If two users are not meeting in a given time frame, the model will prefer to send data through the internet. If they are expected to meet in a given time frame, the system will transfer the data between these users using the D2D approach to minimise energy consumptions.

The results shown in Figure 5.6 come from Algorithm-1 in Figure 5.3; the multiple-criteria decision model will decide the optimal mode of communication between different user's demands. For generating the result, the USC traces of the users are used. The user's data demands are fixed and only contain the user's traces. The results are generated by assuming 200 users. The delay for demands of the data transmitting is considered for day-1, day-2, day-3, day-4, and day-5.

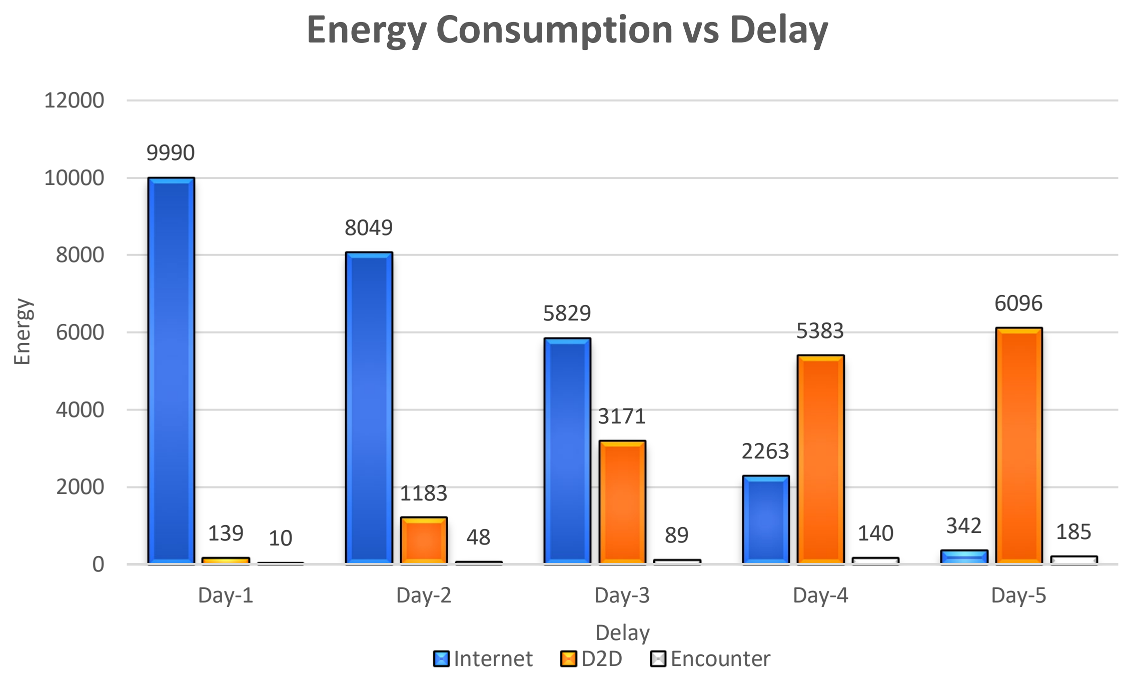


Figure 5.6 Energy Consumption vs Delay

This comparison plot shows as delay increases the more data will transfer through D2D. A set of 200 nodes communicating the data to each other in random data size, based on multiple-criteria decisions to decides the data may transfer through D2D or the internet.

On the 1st day, the ten users seem to appear in the same physical location as each other, transferring the data through D2D communication, and consuming 139 J of energy. The other 190 users are transferring the data through the internet, consuming an energy of 9990 J.

On the 2nd day, the 48 users can appear at the same location and transfer the data through D2D, consuming energy of 1183 J. Other users are transferring through the internet, consuming an energy of 8049 J.

On the 5th day, the 185 users can appear at the exact location and transfer the data through D2D, consuming energy of 6096 J. The other 15 users are transferring the data through the internet, consuming an energy of 342 J.

It shows as delay increases, the more chances of data transferred through D2D. If data is not urgent, the delay option is better to transfer the data and save energy without compromising the QoS.

The user modes are decided based on their traces, so to sustain the QoS, our model EEDDA approach ensures the data delivery through the internet.

Considering the results of this plot, it can conclude that the delay option is efficient as the delay option may transfer the data through D2D and save energy. The results concluded that the multiple-criteria decision model becomes an optimal approach if the severity of the data is low, which can be transmitted in multiple days. More data will be sent through D2D modes, resulting in optimising the energy consumption between the users.

5.4.5. Energy Comparison based on the availability of users at each location.

EEDDA model optimises the power consumption by using the multiple-criteria decision model for the availability of users on a specific location.

For the case study, the five different locations are considered with 200 users on each location who wanted to communicate with each other. After utilising the multiple-criteria decision model on each location, assuming the different percentage of users physically available at the same location such as location-1 has 45% of the total users; total users available at location-1 are 90. The location-2 has 55% of the total users; the total users available at location-2 are 100. Location-3 has 70% of the total users; the total users available at location-3 are 140. location-4 has 85% of the total users; the total number of users available at location-4 is 170; location-5 has 45% of the total users, total users available at location-5 are 190.

Our multiple-criteria decision model will choose D2D communication for the transfer of data based on the available users on each location, and the remaining users will transfer the data through the internet.

The EEDDA decides for the network; for example, if users want to transmit the data at location 1, the model will check the similar user for location-1. These users send the data through D2D communication rest of the user's data will be transferred through the internet.

It is evident from Figure 5.7 that as the number of availability of users increases in the exact location, most data will be transferred through D2D communication and result in reducing the total energy consumption.

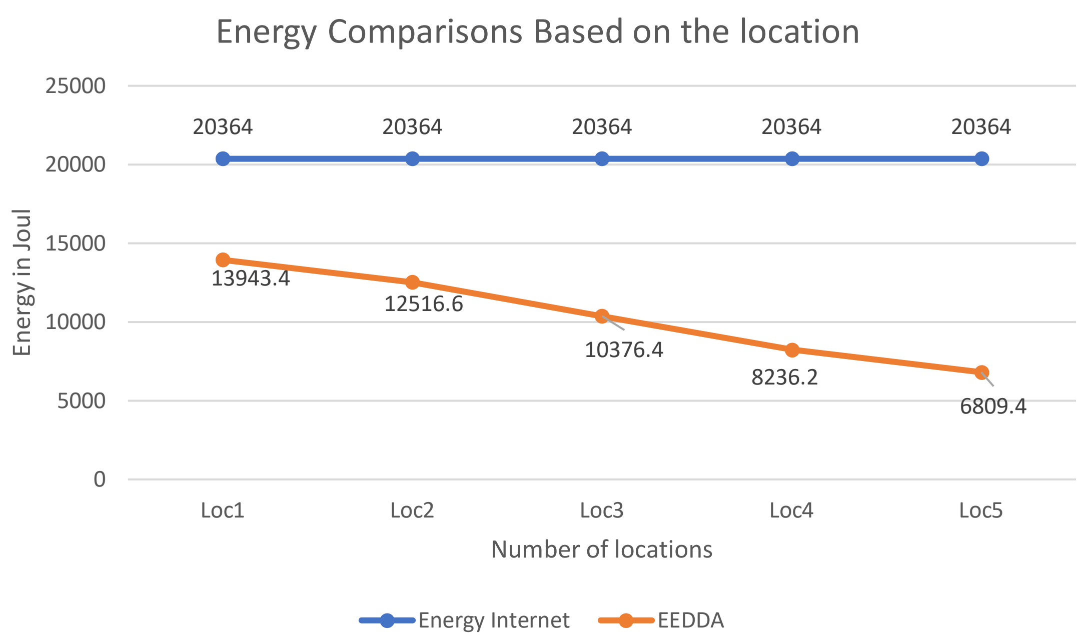


Figure 5.7 Energy consumption vs location

The plot shows that EEDDA is utilising a multiple-criteria decision model for energy consumption as compared to internet communication. The EEDDA results in less energy consumption as compared to the normal flow of communication (internet).

5.5 Energy-Efficient Data Dissemination Approach Privacy and Profiling Issues

Realistically, implementing any D2D communication paradigm need to incorporate the security mechanism for privacy and profiling control mechanism. Several questions need to be addressed to realise efficient EEDDA network architecture. In the first place, EEDDA proposes a network paradigm in which D2D can directly exchange information with delay-tolerant nature to deviate from the infrastructure-based transmission approach. These devices may have limited processing ability to perform cryptographic computation and cryptographic key management; the devices are autonomous and sometimes anonymous and distributed in nature. How can a device know whether its communication partner is genuine or malicious? Can the request, content, or message from any mobile device be trusted? How can the devices perform self-organisation processes to provide a secure network paradigm with a high QoS and absolute privacy control? To fully answer these questions, EEDDA architecture needs to incorporate a mechanism for efficient traffic filtering control. In this line, a distributed algorithmic approach is required for tracking the operating state and characteristics of packet context and network connections for dynamic packet filtering; and data transmission control across devices. EEDDA anticipates a significant privacy control mechanism through developing a distributed context-aware trust algorithm, traffic engineering decisions, and a stateful traffic mechanism for privacy and profiling handling in EEDDA architecture.

Finally, it needs to state that the current proof-of-concept on the EEDDA paradigm does not yet lend itself to sweeping prescription. Nevertheless, this research direction is thought-provoking and paving a new conversation and research direction for researchers to rethink and redesign more sustainable communications and networking by fully exploring the existing infrastructures of telecommunications, transportations and social networking, as well as the mobility and encounter opportunities of the moving objectives, include human and things.

5.6 Summary

A new approach to taking multiple-criteria decisions to reduce energy usage and maintain the standard of service was suggested in this chapter. Our proposed EEDDA approach checks the availability and stability of users and decides for networks.

Our EEDDA decides on a network (D2D or Internet) to transfer data based on the availability of users (location and time).

The work best utilises the existing infrastructures of telecommunications, infrastructure-less, mobile social networking in proximity, to transmit the traffic with delay-tolerant features. The key idea is to reduce the infrastructure-based transmissions by utilising the devices' mobility to encounter stalking among intelligent devices through the short-range D2D tethered communications using the internet of moving things. A case study and a mathematical model are presented to minimise energy consumption. The results show that the proposed model is more efficient in terms of energy consumption. The results are verified for the device to device and the internet.

The case study mentioned in this chapter offers clear evidence that substantial energy savings can be achieved while ensuring data transmission. To create more adaptable data delivery methods, complex mobility models are to be used.

Chaper 6
Conclusion

Smart devices are our lifestyle's digital centre for smooth ways of accomplishing traditional everyday tasks. Such devices have many robust features, including mobility, ease of use, and knowledge of the surrounding environment, which induce consumers to consider them a private asset rather than a digital tools. In addition to the vast number of applications available for work, entertainment, and social networking, mobile devices provide many methods for connecting with the rest of the world, such as cellular networks, Wi-Fi, and ad-hoc mode. Today, mobile devices are now fitted with a rich array of devices and communication interfaces that provide different details about the user's environment and condition, such as their current position and motion. The mobile's microphone, camera, and communication interfaces, including Wi-Fi and Bluetooth, can be utilised as devices to reveal decisive information about the environment nearby.

The purpose of this thesis is to evaluate sustainable next-generation network design. A new advancement for energy efficiency in the opportunistic network is proposed using social awareness and delay-tolerant approaches. In this way, the solution presents that user mobility similarity matrixes can help implement energy-saving while making a more effective decision in forwarding messages. The co-occurrence matrix will help to show the user's mobility pattern or encounter. Based on encounters, it can be analysed that the data will be transferred through D2D, and the energy will be saved. To contribute new knowledge and solutions to prove that this idea is better than traditional approaches to sustain QoS and energy efficiency simultaneously. A proposed network model represents our problem to calculate a more sustainable energy approach.

This thesis proposed utilizing the existing telecommunications infrastructures, infrastructure-less mobile social networking in proximity, to transmit the traffic with delay-tolerant features. Our experiments show that the D2D saves energy consumption with the help of multiple-criteria decisions. An experimental setup is used to analyse this research and obtained results for the multiple-criteria decision energy model.

6.1 Summary of Contributions

Mobile and wireless networks are part of our lives. More and more people are becoming part of interactive networks to exchange multiple details. A Mobile Ad-hoc Network (MANET) is a decentralised network-less infrastructure job in which wireless nodes switch arbitrarily. The nodes communicate directly with nodes in the transmission range and engage with others in multi-hop communication. Such networks are rapidly gaining popularity because of their easy integration. However, due to the rapid economic growth influenced by industrialisation and globalisation, energy consumption has steadily increased in recent years. Industry, transport, and buildings are the three main economic sectors that consume a large amount of energy, having the highest proportion of homes. Energy consumption is a significant problem in mobile ad-hoc wireless networks, as most mobile nodes run on a small battery resource. Current models for measuring energy usage in mobile ad-hoc networks have shown that the different components of energy-related costs involve both transmission power and processing capability.

Device-to-device (D2D) connectivity has been suggested as an underlying coexistence with cellular networks to increase cell throughput and save the device's transmitting capacity. D2D connectivity underlying cellular networks has also gained significant attention recently. The reason for this study, which is the identification of a research gap between the studies in the computer network, similarity analysis, encounter prediction on 'D2D' versus Internet communication, and saves energy consumption from sustaining QoS, has been verified from the literature review in these areas in Chapter 2.

Chapter 3 summarises similarity analysis in which a relationship is developed between distance and node energy use. The distance from data transmission is a key but variable factor contributing to the overall energy usage. It is clear that the longer the period of transmission, the greater the overall energy consumption. Second, it consists of a background that provides a complete summary of human mobility activity features, distorted location visiting preferences, mobility heterogeneity, mobility correlations, IP address, and MAC address. Chapter 3 offers a brief description of trends for similarity analysis. Our proposed approach in this chapter utilises system details, i.e. IP address and MAC folder. The proposed solution assumes that all users have the same accessibility character rather than provide unique data communication networks. The same network helps to display specific mobile behaviours.

A few novel conduct mindful conventions and directions are intended for DTNs, utilising closeness as a foundation for their engineering. In the chapter, the spatial-fleeting conduct closeness profiles among portable clients are examined. When characterising versatility profiles dependent on clients' affiliation frameworks and afterwards utilise indirect similarity method-based-weighted similarity records to look at these versatility profiles quantitatively.

The Similarity Analysis Approach (SAA) is provided for finding behaviour similarity of users' mobility. First, it is shown how our mobile phone can be used as a user's mobility behaviour analyser. Most people carry their mobile phones with them throughout the day; hence it is the most suitable and cost-efficient way of analysing mobility behaviour. The existing mobility modelling approach for data transmission among users is also provided. At first, there is a need to find the similarity among user's behaviour; the spatial-temporal parameters of a mobile user and the device-network information of the mobile node is used. Then, the spatial preferences of a user are provided along with temporal values at which a particular location was visited. The mathematical way of calculating the similarity was provided in the third chapter. For experiments, the mobility traces of users from different campuses, i.e. (campus1, campus2, …, campus6) are used. After extracting the spatial-temporal associations of traces, it is stored in the co-occurrence matrix and normalised the time associations, use the direct similarity method to conclude the similarity. The whole process of calculating behavioural similarity is explained through the case study. In experiments, similarity analysis of the processed data is performed and calculated the similarity of different users on different days and different campuses. In the future, it is the plan to further investigate similarity modelling and prediction and introduce new methods and approaches for similarity analysis.

Second, the description of the prediction of encounters using a random forest model is provided. Chapter 4 compared and Analysed the Gaussian process and random forest using data traces of real-world mobility and showed how these results of encounter predictions save more energy. The precision of expected encounters at different locations is calculated by all methods to fit the products provided by the Gaussian and Random Forest models. Such studies compared the expected encounter precisely at a similar location in different weeks of the month. Also, in this chapter, the Random forest model is used to predict user encounters. The proposed work determines the exact location, and time, at which the encounter will occur with a specific node. The purpose of the encounter prediction chapter was to lessen the resources consumed for continuously searching any available nearby node with which the encounter may or may not occur.

The human mobility patterns for an efficient, ensuring, and less resource (energy) consuming data transmission process in the previous chapter. The encounter prediction is presented, i.e. predicting when two or more users will encounter a specific place and time. A random forest model is used that stored the history of human mobility traces in tree form for encounter prediction. For encounter prediction, the model first predicted the next location-based of mobile users based on the data stored in decision trees in the random forest model. The encounter prediction was made based on data stored in the decision trees. A mathematical representation of the random forest model is also provided that shows how to encounter different users is predicted using the mobility history data of different users. The model showed how the trees in the model grow based on the value of a random variable. The steps for determining specific positions for the cuts in the tree were also explained. The process of working encounter prediction is presented with the help of a scenario. At first, the introduction and the need for an energy-efficient data dissemination approach using multiple-criteria decisions are mentioned. Then some existing results and techniques for energy-efficient data dissemination are listed. The model used for energy-efficient data dissemination is explained along with the mathematical representation of the model. The experiments, results, and comparison with standard techniques are also discussed in detail at the end.

In the option of data transmission through the Internet, energy usage depends on the energy used by nodes at both ends for transmitting and receiving sides to download and upload and the amount of network equipment used, such as hops used in between before the final destination. Data transmission through D2D energy consumption depends upon data volume, sending, and receiving sides. Our goal is to find out which communication way is using less energy consumption. In chapter 5, multiple-criteria decision energy models are used. The model shows the multiple-criteria decisions, and based on encounter prediction results, the node decides to transfer the data through the Internet or D2D. The number of recent researches addressing D2D communications [2] and their implementations is wide. It ranges from those addressing offloading mobile data traffic to those dealing with the extension of network coverage or sharing of content in community areas.

The human mobility patterns are used for an efficient, ensuring, and less resource (energy, bandwidth) consuming data Transmission process. The proposed model knows energy use about occupant behavioural trends from the building. In the encounter prediction, i.e. predicting when two or more users will encounter at a specific place and time. The exact encounter prediction can help avoid resource consumption used continuously searching for any available encounter. The random forest model is used to store the history of human mobility traces in tree form for encounter prediction. For encounter prediction, the model at first predicts the next future location. Then based on similar mobility preferences, an encounter is predicted at a specific day of the month and hour of the day. A mathematical model is provided to encounter the prediction of different users. For experimentation, The USC dataset is trained with Benchmark Gaussian and Random forest model. The results were computed for observed and predicted encounters through both approaches. The accuracy of predicted encounters and their comparison were also shown in chapter 5. The results showed that the accuracy of encounter prediction through the random forest model was relatively higher than the Benchmark Gaussian model. Even at some locations, the encounters were predicted even with accuracy higher than 80%. The Gaussian process and random forest are compared and analysed using real-world mobility data traces and show how these encounter prediction results save around 30% of the energy used compared to other communication models and show that device to device (D2D) are better than the internet. Our proposed method can save energy using the Random Forest Model. In the future, it is the plan to investigate energy-saving models.

Human versatility designs are beneficial in information transmission measures among various hubs. The portability designs help to characterise client conduct as development. These versatility designs help to predict the future area of a client. The information on the future area of clients is significant for inducing different boundaries, for example, next future experience. This section has utilised human portability designs for a productive, guaranteeing, and less asset (energy, transfer speed) burning-through information transmission measure. It is introduced to deal with experience forecast, for example, foreseeing when at least two clients will experience at a particular spot and time. The specific experience forecast can assist with evading asset utilisation that is utilised consistently, looking for any accessible experience. For experience forecast, an irregular backwoods model is utilised to put away the historical backdrop of human versatility in a tree structure. For experience forecast, the model anticipated the following future area put together of portable clients based on the information put away in choice trees in the arbitrary backwoods model. In light of the comparative portability inclinations of various clients, the experience was anticipated at a particular day and hour of the day. The experience forecast was based on information put away in the choice trees. A numerical portrayal of an arbitrary woodland model is given that shows how to experience among various clients is anticipated utilising versatility history information of multiple clients.

The USC dataset with Benchmark Gaussian and Random woods model is prepared for experimentation. The climate for tests was created in Python. From the outset, the watched experiences were noted down. For results, the experiences were anticipated through both, Random timberland and Gaussian approaches. A cross relationship was also plotted through the two methodologies for results examination. The correctness of anticipated experiences through the two methodologies, their graphical portrayal, and correlation have additionally appeared in part. The outcomes indicated that the precision of experience expectation through the irregular old models was lower than the Benchmark Gaussian model. Indeed, even in certain areas, the experiences were predicted even with accuracy higher than 80%.

Thirdly, uses several decisions to present an energy-efficient data dissemination approach was presented in chapter 5. This chapter will address the challenges of energy first and then how the EEDDA approach checks the encounter prediction results and stability of encounters and make multiple-criteria decisions.

System Model for Energy Efficient Data Dissemination Approach is proposed in chapter 5. Then, explained the transmission of data through an internet network. The device-internet and device-device energy usage are analysed. In the proposed approach, EEDDA sustains the QoS (Quality of Service).

Fourthly, this proposed approach takes multiple-criteria decisions to reduce energy consumption and sustain the Quality of service. In chapter 5 EEDDA take the results of chapter 4 encounter prediction based on prediction results to check some specific condition such as encounter time, data size, delay, etc. and based on these take the decision the data goes to the internet of the device to device.

Our proposed EEDDA approach checks the encounter prediction results and stability of encounters and takes multiple-criteria decisions. The work is the best utilisation of the existing telecommunications infrastructures, infrastructure-less mobile social networking in proximity, to transmit the traffic with delay-tolerant features. The key idea is to reduce the infrastructure-based transmissions by utilising the devices' mobility to encounter stalking among intelligent devices through the short-range Device-to-Device (D2D) tethered communications using the Internet of moving things. As the last chapter 5, this thesis also presented a case study and a mathematical model for minimised energy consumption. The results show that the proposed model is more efficient in terms of energy consumption. The results are verified for the device to device and the internet. Clear evidence is shown that substantial energy savings can be achieved while ensuring data delivery with the help of the case study. The work has contributed to the research by presenting a model for multiple-criteria decision energy consumption among different users having mobile phones. The proposed model will show the numerous decision of node takes the decision based on encounter prediction results in data transfer through the D2D or Internet.

The 'novelty of the work' lies in the reduced resource consumption that our proposed model offers. The key idea is to minimise the infrastructure-based transmissions by utilising the devices' mobility to encounter stalking among smart devices through the short-range Device-to-Device (D2D) tethered communications using the Internet of moving things. Based on the similarity analysis of the real mobility traces, this thesis has confirmed the rich opportunities for mobile encounters in human society. Hence, in this research, the proposed model EEDDA reduces energy consumption and sustains the Quality of Services.

The thesis proposed the best utilisation of the existing telecommunications infrastructures, infrastructure-less mobile social networking in proximity, to transmit the traffic with delay-tolerant features. The results of our experiments show the D2D saves energy consumption with the help of multiple-criteria decisions. This thesis addressed the energy challenges first and then gave models and approaches to overcome them.

6.2 Future Work

The study of the impact of the proposed solution on D2D communication and the study of the energy consumption introduced by this approach remains part of our ongoing and future research work.

Considering the work covered in this thesis occurred within the constraint of a limited period and the expected development of the future network, it would be helpful to highlight some future areas to be further investigated.

Future work evaluates the proposed work against scalability, evaluates different data communication problems, and applies the algorithm for different attacks during data transfer.

First, to build more adaptable data distribution approaches, it is expected to use the complex mobility models and consider some unpredictability of public transport movements. Secondly, there should be more approaches for similarity that must be introduced for different scenarios. In the future, the researchers may propose more models that may help reduce energy consumption using the encounter prediction method. Finally, more research is needed to combine and integrate some of the approaches presented in this thesis to keep better communication functioning for a longer duration.

References

[1] E. Borgia, D. G. Gomes, B. Lagesse, R. Lea, and D. Puccinelli, "Special issue on Internet of Things: Research challenges and Solutions," *Computer Communications,* vol. 89, pp. 1-4, 2016.

[2] S. R. Bhandari, G. M. Lee, and N. Crespi, "Mobility Model for User's Realistic Behavior in Mobile Ad Hoc Network," in *8th Annual Communication Networks and Services Research Conference*, 2010: IEEE, pp. 102-107.

[3] A. Ali *et al.*, "The MGB-5 Challenge: Recognition and Dialect Identification of Dialectal Arabic Speech," in *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019: IEEE, pp. 1026-1033.

[4] K. K. Jahromi, M. Zignani, S. Gaito, and G. P. Rossi, "Simulating human mobility patterns in urban areas," *Simulation Modelling Practice and Theory,* vol. 62, pp. 137-156, 2016.

[5] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer networks,* vol. 54, no. 15, pp. 2787-2805, 2010.

[6] S. Li, L. Da Xu, and S. Zhao, "The internet of things: a survey," *Information Systems Frontiers,* vol. 17, no. 2, pp. 243-259, 2015.

[7] K. K. Patel and S. M. Patel, "Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges," *International Journal of Engineering Science and Computing,* vol. 6, no. 5, 2016.

[8] L. Cheng, S. R. Geedipally, and D. Lord, "The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data," *Safety Science,* vol. 54, pp. 38-42, 2013.

[9] P. Juliana *et al.*, "Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics," *Nature Genetics,* vol. 51, no. 10, pp. 1530-1539, 2019.

[10] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems,* vol. 29, no. 7, pp. 1645-1660, 2013.

[11] T. Chen, Y. Yang, H. Zhang, H. Kim, and K. Horneman, "Network Energy Saving Technologies for Green Wireless Access Networks," *IEEE Wireless Communications,* vol. 18, no. 5, pp. 30-38, 2011.

[12] K. Khalid, I. Woungang, S. K. Dhurandher, L. Barolli, G. H. Carvalho, and M. Takizawa, "An Energy-Efficient Routing Protocol for Infrastructure-less

Opportunistic Networks," in *10th international Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, 2016: IEEE, pp. 237-244.

[13] R. K. Ganti, F. Ye, and H. Lei, "Mobile Crowdsensing: Current State and Future Challenges," *IEEE Communications Magazine,* vol. 49, no. 11, pp. 32-39, 2011.

[14] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and Big Data Analytics for Smart and Connected Communities," *IEEE access,* vol. 4, pp. 766-773, 2016.

[15] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: A New Mobility Model for Human Walks," in *IEEE INFOCOM 2009*, 2009: IEEE, pp. 855-863.

[16] P. Levis, T. Clausen, J. Hui, O. Gnawali, and J. Ko, "The Trickle Algorithm," *Internet Engineering Task Force, RFC6206,* 2011.

[17] S. Mclaughlin *et al.*, "Techniques for Improving Cellular Radio Base Station Energy Efficiency," *IEEE Wireless Communications,* vol. 18, no. 5, pp. 10-17, 2011.

[18] A. Taha, R. Alsaqour, M. Uddin, M. Abdelhaq, and T. Saba, "Energy Efficient Multipath Routing Protocol for Mobile Ad-Hoc Network Using the Fitness Function," *IEEE Access,* vol. 5, pp. 10369-10381, 2017.

[19] S. Sarkar and R. Datta, "A secure and energy-efficient stochastic multipath routing for self-organized mobile ad hoc networks," *Ad Hoc Networks,* vol. 37, pp. 209-227, 2016.

[20] J. Hoebeke, I. Moerman, B. Dhoedt, and P. Demeester, "Adaptive Multi-mode Routing in Mobile Ad Hoc Networks," in *IFIP International Conference on Personal Wireless Communications*, 2004: Springer, pp. 107-117.

[21] Y. Hao, L. Haiyun, Y. Fan, L. Songwu, and Z. Lixia, "Security in Mobile Ad Hoc Networks: Challenges and Solutions," *IEEE Wireless Communications,* vol. 11, no. 1, pp. 38-47, 2004.

[22] A. P. Bianzino, C. Chaudet, D. Rossi, and J.-L. Rougier, "A Survey of Green Networking Research," *IEEE Communications Surveys & Tutorials,* vol. 14, no. 1, pp. 3-20, 2010.

[23] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," in *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2003, pp. 27-34.

[24] Z. Zhang, "Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges," *IEEE Communications Surveys & Tutorials,* vol. 8, no. 1, pp. 24-37, 2006.

[25] J. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater sensor networks: applications, advances and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 370, no. 1958, pp. 158-175, 2012.

[26] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic Networking: Data Forwarding in Disconnected Mobile Ad Hoc Networks," *IEEE communications Magazine,* vol. 44, no. 11, pp. 134-141, 2006.

[27] M. Conti and S. Giordano, "Mobile Ad Hoc Networking: Milestones, Challenges, and New Research Directions," *IEEE Communications Magazine,* vol. 52, no. 1, pp. 85-96, 2014.

[28] A. Shaikh, D. Vasan, and K. H. Mohammadani, "Performance Analysis of MANET Routing Protocols-A Comparative Study," *International Journal of Computer Applications,* vol. 83, no. 7, 2013.

[29] A. Nguyen and N. Crespi, "SOCIAL-DTN: Why Social Networking Services is more fruitful to Mobile Delay-Tolerant Networks?," in *2009 International Conference on Ultra Modern Telecommunications & Workshops*, 2009, pp. 1-2.

[30] F. Nazir, H. Prendinger, and A. Seneviratne, "Participatory Mobile Social Network Simulation Environment," in *2010 IEEE International Conference on Communications*, 2010: IEEE, pp. 1-6.

[31] A. Liu, Y. Xiao, H. Lu, S.-B. Tsai, and W. Song, "A fuzzy three-stage multi-attribute decision-making approach based on customer needs for sustainable supplier selection," *Journal of Cleaner Production,* vol. 239, p. 118043, 2019.

[32] A. Ghani, M. Zubair, S. A. K. Ghayyure, and S. Muhammad, "Honesty Based Democratic Scheme to Improve Community Cooperation for IoT Based Vehicular Delay Tolerant Networks," *arXiv e-prints,* p. arXiv: 2004.08698, 2020.

[33] A. Lindgren, A. Doria, E. Davies, and S. Grasic, "Probabilistic Routing Protocol for Intermittently Connected Networks," *RFC Series,* 2012.

[34] C. Boldrini, M. Conti, J. Jacopini, and A. Passarella, "Hibop: a History Based Routing Protocol for Opportunistic Networks," in *2007 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2007: IEEE, pp. 1-12.

[35] A. H. Sodhro, S. Pirbhulal, Z. Luo, and V. H. C. De Albuquerque, "Towards an optimal resource management for IoT based Green and sustainable smart cities," *Journal of Cleaner Production,* vol. 220, pp. 1167-1179, 2019.

[36] F. Bai and B. Krishnamachari, "Spatio-Temporal Variations of Vehicle Traffic in VANETs: Facts and implications," in *VANET '09*, 2009.

[37] D. Kotz and T. Henderson, "Crawdad: A Community Resource for Archiving Wireless Data at Dartmouth," *IEEE Pervasive Computing,* vol. 4, no. 4, pp. 12-14, 2005.

[38] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási, "Uncovering individual and collective human dynamics from mobile phone records," *Journal of physics A: Mathematical and Theoretical,* vol. 41, no. 22, p. 224015, 2008.

[39] F. Ekman, A. Keränen, J. Karvo, and J. Ott, "Working Day Movement Model," in *Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models*, 2008, pp. 33-40.

[40] L. Pappalardo and F. Simini, "Data-driven generation of spatio-temporal routines in human mobility," *Data Mining Knowledge Discovery,* vol. 32, no. 3, pp. 787-829, 2018.

[41] G. S. Thakur, A. Helmy, and W.-J. Hsu, "Similarity Analysis and Modeling in Mobile Societies: the Missing Link," in *Proceedings of the 5th ACM workshop on Challenged networks*, 2010, pp. 13-20.

[42] M. ElSherief, T. ElBatt, A. Zahran, and A. Helmy, "The Quest for User Similarity in Mobile Societies," in *IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS)*, 2014: IEEE, pp. 569-574.

[43] J. Lin, B. Liu, N. Sadeh, and J. I. Hong, "Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings," in *10th Symposium On Usable Privacy and Security {SOUPS}*, 2014, pp. 199-212.

[44] M. Lv, L. Chen, and G. Chen, "Mining user similarity based on routine activities," *Information Sciences,* vol. 236, pp. 17-32, 2013.

[45] N. Akhtar Khan, F. Ahmad, and S. A. Khan, "Formal and Executable Specification of Random Waypoint Mobility Model Using Timed Coloured Petri Nets for WMN," *Abstract and Applied Analysis,* vol. 2014, p. 798927, 2014.

[46] V. C. Barbosa, R. Donangelo, and S. R. Souza, "Emergence of Scale-Free Networks from Local Connectivity and Communication Trade-offs," *Physical Review E,* vol. 74, no. 1 Pt 2, p. 016113, 2006.

[47] A. Reynolds, "Truncated Lévy walks are expected beyond the scale of data collection when correlated random walks embody observed movement patterns," *Journal of the Royal Society Interface,* vol. 9, no. 68, pp. 528-534, 2012.

[48] S. Hong, K. Lee, and I. Rhee, "STEP: A Spatio-Temporal Mobility Model for Humans Walks," in *The 7th IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE MASS 2010)*, 2010: IEEE, pp. 630-635.

[49] A. Munjal, T. Camp, and W. C. Navidi, "SMOOTH: A Simple Way to Model Human Walks," *ACM SIGMOBILE Mobile Computing and Communications Review,* vol. 14, no. 4, pp. 34-36, 2010.

[50] J. Wang, X. Kong, F. Xia, and L. Sun, "Urban Human Mobility: Data-Driven Modeling and Prediction," *Acm Sigkdd Explorations Newsletter,* vol. 21, no. 1, pp. 1-19, 2019.

[51] B. F. Nielsen, K. Sneppen, L. Simonsen, and J. Mathiesen, "Social network heterogeneity is essential for contact tracing," *medRxiv,* 2020.

[52] N. Fukae, A. Fujihara, and H. Miwa, "Model for Generation of Social Network Considering Human Mobility," in *International Conference on Intelligent Networking and Collaborative Systems*, 2018: Springer, pp. 294-305.

[53] J. Singh, S. K. Dhurandher, and V. Kumar, "Mobility Models in Opportunistic Networks," in *Opportunistic Networks*: CRC Press, 2021, pp. 225-242.

[54] O. S. Younes and U. A. Albalawi, "Analysis of Route Stability in Mobile Multihop Networks Under Random Waypoint Mobility," *IEEE Access,* vol. 8, pp. 168121-168136, 2020.

[55] Z. Yao, "Exploiting Human Mobility Patterns for Point-of-Interest Recommendation," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 2018, pp. 757-758.

[56] J. Ryu, J. Park, J. Lee, and S.-B. Yang, "Community-based diffusion scheme using Markov chain and spectral clustering for mobile social networks," *Wireless Networks,* vol. 25, no. 2, pp. 875-887, 2019.

[57] F. Zhou, R. Yin, G. Trajcevski, K. Zhang, J. Wu, and A. Khokhar, "Improving human mobility identification with trajectory augmentation," *GeoInformatica,* vol. 25, no. 3, pp. 453-483, 2021.

[58] J. Feng, Z. Yang, F. Xu, H. Yu, M. Wang, and Y. Li, "Learning to Simulate Human Mobility," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3426-3433.

[59] H. Shi, Y. Li, H. Cao, X. Zhou, C. Zhang, and V. Kostakos, "Semantics-Aware Hidden Markov Model for Human Mobility," *IEEE Transactions on Knowledge and Data Engineering,* 2019.

[60] P. Mazumdar, "Human Mobility Analysis in Location-based Social Networks," 2018.

[61] E. M. Daly and M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-tolerant Manets," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2007, pp. 32-40.

[62] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of Human Mobility on Opportunistic Forwarding Algorithms," *IEEE Transactions on Mobile Computing,* vol. 6, no. 6, pp. 606-620, 2007.

[63] W.-j. Hsu, D. Dutta, and A. Helmy, "CSI: A paradigm for behavior-oriented profile-cast services in mobile networks," *Ad Hoc Networks,* vol. 10, no. 8, pp. 1586-1602, 2012.

[64] W. Glänzel and B. Thijs, "Using 'core documents' for the representation of clusters and topics," *Scientometrics,* vol. 88, no. 1, pp. 297-309, 2011.

[65] B. Jarneving, "A variation of the calculation of the first author cocitation strength in author cocitation analysis," *Scientometrics,* vol. 77, no. 3, pp. 485-504, 2008.

[66] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a Location Predictor on Trajectory Pattern Mining," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 637-646.

[67] L. Vu, Q. Do, and K. Nahrstedt, "Jyotish: A Novel Framework for Constructing Predictive Model of People Movement from Joint Wifi/Bluetooth Trace," in *2011 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2011: IEEE, pp. 54-62.

[68] H. H. Bauer, T. Reichardt, S. J. Barnes, and M. M. Neumann, "Driving Consumer Acceptance of Mobile Marketing: A Theoretical Framework and Empirical study," *Journal of Electronic Commerce Research,* vol. 6, no. 3, p. 181, 2005.

[69] D. Lymberopoulos, A. Bamis, and A. Savvides, "Extracting Spatiotemporal Human Activity Patterns in Assisted Living using a Home Sensor Network," in *Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments*, 2008, pp. 1-8.

[70] J. Li *et al.*, "Intelligent identification of effective reservoirs based on the random forest classification model," *Journal of Hydrology,* vol. 591, p. 125324, 2020.

[71] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, "Random Forest based hourly building energy prediction," *Energy and Buildings,* vol. 171, pp. 11-25, 2018.

[72] L. Breiman, "Bagging Predictors," *Machine Learning,* vol. 24, no. 2, pp. 123-140, 1996.

[73] L. Breiman, "Some Infinity Theory for Predictor Ensembles," Citeseer, 2000.

[74] R. Genuer, J.-M. Poggi, and C. Tuleau, "Random Forests: some methodological insights," *arXiv preprint arXiv:0811.3619,* 2008.

[75] G. U. Rehman, A. Ghani, S. Muhammad, M. Singh, and D. Singh, "Selfishness in Vehicular Delay-Tolerant Networks: A Review," *Sensors,* vol. 20, no. 10, p. 3000, 2020.

[76] P. Jiang, H. Wu, W. Wang, W. Ma, X. Sun, and Z. Lu, "MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features," *Nucleic Acids Research,* vol. 35, no. suppl_2, pp. W339-W344, 2007.

[77] W. Chen *et al.*, "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility," *Catena,* vol. 151, pp. 147-160, 2017.

[78] C. Williams, E. V. Bonilla, and K. M. Chai, "Multi-task Gaussian Process Prediction," *Advances in Neural Information Processing Systems,* pp. 153-160, 2007.

[79] S. Shahbazi, S. Karunasekera, and A. Harwood, "Improving Performance in Delay/Disruption Tolerant Networks through Passive Relay Points," *Wireless Networks,* vol. 18, no. 1, pp. 9-31, 2012.

[80] C. Catal, U. Sevim, and B. Diri, "Clustering and Metrics Thresholds Based Software Fault Prediction of Unlabeled Program Modules," in *Sixth International Conference on Information Technology: New Generations*, 2009: IEEE, pp. 199-204.

[81] T.-K. Huang, C.-K. Lee, and L.-J. Chen, "Prophet+: An Adaptive Prophet-Based Routing Protocol for Opportunistic Network," in *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, 2010: IEEE, pp. 112-119.

[82] M. M. Tajiki, S. Salsano, L. Chiaraviglio, M. Shojafar, and B. Akbari, "Joint Energy Efficient and QoS-Aware Path Allocation and VNF Placement for Service Function Chaining," *IEEE Transactions on Network and Service Management,* vol. 16, no. 1, pp. 374-388, 2019.

[83] M. A. Marsan and M. Meo, "Green Wireless Networking: Three Questions," in *The 10th IFIP Annual Mediterranean Ad Hoc Networking Workshop*, 2011: IEEE, pp. 41-44.

[84] E. Yaacoub, H. Ghazzai, M.-S. Alouini, and A. Abu-Dayya, "Achieving Energy Efficiency in LTE with Joint D2D Communications and Green Networking Techniques," in *9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2013: IEEE, pp. 270-275.

[85] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, "Four Degrees of Separation, Really," in *Proceedings of the 4th Annual ACM Web Science Conference*, 2012, pp. 33-42.

[86] K. Massri, A. Vitaletti, A. Vernata, and I. Chatzigiannakis, "Routing Protocols for Delay Tolerant Networks: A Reference Architecture and a Thorough Quantitative Evaluation," *Journal of Sensor and Actuator Networks,* vol. 5, no. 2, p. 6, 2016.

[87] G. S. Thakur, R. Tiwari, M. Thai, S.-S. Chen, and A. Dress, "Detection of local community structures in complex dynamic networks with random walks," *IET Systems Biology,* vol. 3, no. 4, pp. 266-278, 2009.

[88] J. Makhlouta, H. Harkous, F. Hutayt, and H. Artail, "Adaptive Fuzzy Spray and Wait: Efficient Routing for Opportunistic Networks," in *International Conference on Selected Topics in Mobile and Wireless Networking (iCOST)*, 2011: IEEE, pp. 64-69.

[89] M. Doshi and A. Kamdar, "Multi-Constraint QoS Disjoint Multipath Routing in SDN," in *Moscow Workshop on Electronic and Networking Technologies (MWENT)*, 2018, pp. 1-5.

[90] R. E. Goldsmith, "Explaining and Predicting Consumer Intention to Purchase Over the Internet: an Exploratory Study," *Journal of Marketing Theory and Practice,* vol. 10, no. 2, pp. 22-28, 2002.

[91] S. Ogawa and F. T. Piller, "Reducing The Risks of New Product Development," *MIT Sloan Management Review,* vol. 47, no. 2, p. 65, 2006.

[92] S. Marini *et al.*, "A Dynamic Bayesian Network model for long-term simulation of clinical complications in type 1 diabetes," *Journal of Biomedical Informatics,* vol. 57, pp. 369-376, 2015.

[93] P. Ghosekar, G. Katkar, and P. Ghorpade, "Mobile ad hoc networking: imperatives and challenges," *IJCA Special issue on MANETs,* vol. 3, pp. 153-158, 2010.

[94] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Human Behavior Modeling with Maximum Entropy Inverse Optimal Control," in *AAAI Spring Symposium: Human Behavior Modeling*, 2009, vol. 92.

[95] K. C. Lee and H. Cho, "Performance of Ensemble Classifier for Location Prediction Task: Emphasis on Markov Blanket Perspective," *International Journal of u-and e-Service, Science and Technology,* vol. 3, no. 3, p. 2010, 2010.

[96] J. N. Liu, B. N. Li, and T. S. Dillon, "An Improved Naive Bayesian Classifier Technique Coupled with a Novel Input Solution Method," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 31, no. 2, pp. 249-256, 2001.

[97] Q. Wang and J. E. Taylor, "Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms," *Journal of Computing in Civil Engineering,* vol. 30, no. 2, p. 04015004, 2016.

[98] M. Barthélemy, "Spatial networks," *Physics Reports,* vol. 499, no. 1, pp. 1-101, 2011.

[99] L. Pappalardo and F. Simini, "Data-driven generation of spatio-temporal routines in human mobility," *Data Mining and Knowledge Discovery,* vol. 32, no. 3, pp. 787-829, 2018/05/01 2018, doi: 10.1007/s10618-017-0548-4.

[100] O. Babatunde and O. Al-Debagy, "A comparative review of internet protocol version 4 (ipv4) and internet protocol version 6 (ipv6)," *arXiv preprint arXiv:1407.2717,* 2014.

[101] W.-j. Hsu and A. Helmy, "On Nodal Encounter Patterns in Wireless LAN Traces," *IEEE Transactions on Mobile Computing,* vol. 9, no. 11, pp. 1563-1577, 2010.

[102] W.-j. Hsu, D. Dutta, and A. Helmy, "CSI: A Paradigm for Behavior-oriented Delivery Services in Mobile Human Networks," vol. 6, 2008.

[103] W.-j. Hsu, K. Merchant, H.-w. Shu, C.-h. Hsu, and A. Helmy, "Weighted Waypoint Mobility Model and its Impact on Ad Hoc Networks," *ACM SIGMOBILE Mobile Computing and Communications Review,* vol. 9, pp. 59-63, 2005.

[104] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling Spatial and Temporal Dependencies of User Mobility in Wireless Mobile Networks," *IEEE/ACM Transactions on Networking,* vol. 17, no. 5, pp. 1564-1577, 2009.

[105] P. Hui and J. Crowcroft, "Human mobility models and opportunistic communications system design," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences,* vol. 366, no. 1872, pp. 2005-2016, 2008.

[106] W.-j. Hsu, D. Dutta, and A. Helmy, "Mining Behavioral Groups in Large Wireless LANs," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, 2007, pp. 338-341.

[107] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks," *IEEE Transactions on Mobile Computing,* vol. 10, no. 11, pp. 1576-1589, 2010.

[108] M. Kim and D. Kotz, "Periodic properties of user mobility and access-point popularity," *Personal and Ubiquitous Computing,* vol. 11, no. 6, pp. 465-479, 2007.

[109] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment Framework for Participatory Sensing Data Collections," in *International Conference on Pervasive Computing*, 2010: Springer, pp. 138-155.

[110] K. Shilton *et al.*, "Participatory Design of Sensing Networks: Strengths and Challenges," 2008.

[111] H. Kim, M. Bae, W. Lee, and H. Kim, "Adaptive Decision of Wireless Access Network for Higher User Satisfaction," *Wireless Communications and Mobile Computing,* 2018.

[112] C. Chilipirea, A.-C. Petre, and C. Dobre, "Predicting Encounters in Opportunistic Networks using Gaussian Process," in *19th International Conference on Control Systems and Computer Science*, 2013: IEEE, pp. 99-105.

[113] C. Pearson and Z. Hussain, "Smartphone Use, Addiction, Narcissism, and Personality," (in en), *International Journal of Cyber Behavior, Psychology and Learning,* vol. 5, no. 1, pp. 17-32, 2015.

[114] D. Rosenbaum, F. Kurz, U. Thomas, S. Suri, and P. Reinartz, "Towards automatic near real-time traffic monitoring with an airborne wide angle camera system," *European Transport Research Review,* vol. 1, no. 1, pp. 11-21, 2009.

[115] S. S. Chaudhari and R. C. Biradar, "Traffic and mobility aware resource prediction using cognitive agent in mobile ad hoc networks," *Journal of Network and Computer Applications,* vol. 72, pp. 87-103, 2016.

[116] R. Chellapa, A. Jennings, and N. Shenoy, "A Comparative Study of Mobility Prediction in Fixed Wireless Networks and Mobile Ad Hoc Networks," in *IEEE International Conference on Communications,. (ICC '03).* 2003, vol. 2, pp. 891-895.

[117] P. Hui and J. Crowcroft, "Predictability of Human Mobility and Its Impact on Forwarding," *Third International Conference on Communications and Networking in China,* pp. 543-547, 2008.

[118] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: Social-Based Forwarding in Delay-Tolerant Networks," *IEEE Transactions on Mobile Computing,* vol. 10, no. 11, pp. 1576-1589, 2011.

[119] R. I. Ciobanu and C. Dobre, "Predicting Encounters in Opportunistic Networks," in *Proceedings of the 1st ACM workshop on High performance mobile opportunistic systems*, 2012, pp. 9-14.

[120] R. B. Parikh *et al.*, "Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer," *JAMA Network Open,* vol. 2, no. 10, pp. e1915997-e1915997, 2019.

[121] G. Y. Li *et al.*, "Energy-Efficient Wireless Communications: Tutorial, Survey, and Open Issues," *IEEE Wireless Communications,* vol. 18, no. 6, pp. 28-35, 2011.

[122] Q. Qin, Q.-G. Wang, J. Li, and S. S. Ge, "Linear and nonlinear trading models with gradient boosted random forests and application to Singapore stock market," *Journal of Intelligent Learning Systems and Applications,* vol. 5, pp. 1-10, 2013.

[123] C. Yu, B. Lee, and H. Yong Youn, "Energy efficient routing protocols for mobile ad hoc networks," *Wireless Communications and Mobile Computing,* vol. 3, no. 8, pp. 959-973, 2003.

[124] J. Abualdenien, "Proximity Reasoning via Multimodal Context Fusion," 2017.

[125] Z. Abbas and W. Yoon, "A Survey on Energy Conserving Mechanisms for the Internet of Things: Wireless Networking Aspects," *Sensors,* vol. 15, no. 10, pp. 24818-24847, 2015.

[126] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Gódor, "Reducing Energy Consumption in LTE with Cell DTX," in *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011: IEEE, pp. 1-5.

[127] J.-M. Liang, J.-J. Chen, H.-H. Cheng, and Y.-C. Tseng, "An Energy-Efficient Sleep Scheduling with QoS Consideration in 3GPP LTE-Advanced Networks for Internet of Things," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,* vol. 3, no. 1, pp. 13-22, 2013.

Appendix: Published Papers

Paper 1. Ambreen Memon, William Liu, and Adnan Al-Anbuky" CatchMe If You Can: Enable Sustainable Communications Using Internet of Movable Things" *2016 IEEE 14th Intl Conference on Dependable, Autonomic and Secure Computing, 14th Intl Conference on Pervasive Intelligence and Computing, 2nd Intl Conference on Big Data Intelligence and Computing and Cyber Science and Technology.*

Paper 2. Ambreen Memon, William Liu, and Adnan Al-Anbuky" A New Energy Efficient Big Data Dissemination Approach Using the Opportunistic D2D Communications" *Third International Conference, SmartGIFT* 2018 Auckland, New Zealand, April 23–24, 2018.

Paper 3. Ambreen Memon, Jeff Kilby and M.Ishtiaq "Sustainable Smart Connected Buildings: Random Forest-based Energy Saving Model *Elsevier Securing IoT-based Critical Infrastructure Journal* 2021 impact factor 2.6.

Paper 1

2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress

CatchMe If You Can: Enable Sustainable Communications Using Internet of Movable Things

Ambreen Memon, William Liu and Adnan Al-Anbuky
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Emails: {amemon, william.liu and aalanbuk}@aut.ac.nz

Abstract—the customary perspective of Internet of Things (IoT) endeavors to unite all the physical objects or "things" embedded with electronics, software, sensors, and network connectivity to allow more direct integrations between the physical world and cyber-based systems. While these networked devices and associated communications can increase the energy demands exponentially and end up harming the environment seriously. Currently most researchers and practitioners have dedicated their efforts on improving the resource efficiency of IoT systems itself, while missing that there is a great potential to fully utilize the advanced capabilities of storage and communications in IoT devices, especially their movability which can carry the data to the destinations too. In this paper, we are paving a novel communications paradigm, named as **C**ontext-**a**ware **t**ethering **c**ommunications **h**angout using **M**obile **e**ncounter (CatchMe). A time-space matrix has been proposed to capture human mobility preferences, as well as a similarity analytical framework has been developed for mobile encountering analysis and predication, which are pillar stones of CatchMe paradigm. It aims to fully utilize the devices' mobility to create their encounter opportunities for short-range device to device (D2D) communications. Based on the real mobility traces, two case studies for mobility analysis have been conducted and the results confirm that there are reasonable opportunities existing for mobile encountering among most users, thus CatchMe is feasible. The extra numerical study demonstrates that CatchMe has great promises to significantly reduce the energy consumption, comparing with the infrastructure-based transmission approach which has been taken for granted in daily use. The current stage of proof-of-concept on CatchMe paradigm does not yet lend itself to sweeping prescription. Nevertheless, we believe this research direction is thought-provoking and opening a new conversation for researchers to rethink and redesign more sustainable communications and networking by fully exploring the moveable IoTs.

Keywords— Internet of Things; Mobility; Opportunistic routing; Delay tolerant network; Sustainable communications; Device-to-Device (D2D) communications; Similarity analysis

I. BACKGROUND

The Internet of Things (IoT) alludes to the developing pattern of expanding physical objects and gadgets with sensing, detecting, computing and communications correspondence abilities, connecting them to form a system and making utilization of the aggregated impact of the organized objects for improved efficiency, accuracy and economic benefits for human beings. Under the vision of IoT, the cutting edge Internet is to advance the symphonious collaboration between human, social orders, and shrewd things [1]. In recent years, huge examination endeavors have been made on IoT, primarily from a thing-focused point of view. A wide scope of regions are covered including object recognizable proof and following, article systems administration, detecting information perception and protection control etc. [2]. The congruous association in the middle of human and IoT, in other words, the social aspect of IoT, has yet little been investigated.

As far as the network topology highlights, we can comprehensively classify system association into two sorts: infrastructure empowered association, and also specially appointed or astute association of infrastructure-less based. The earlier sort uses the existing infrastructure e.g., mobile base stations, access point, switches in [3-4], while the later utilizes the infrastructure free, short-run radio systems e.g., Bluetooth, Wi-Fi, and so forth, to fabricate decentralized and commercial ad hoc systems. The shrewd systems are human-driven in the light of the fact that they innately take after the way that individuals sharply get into contact. For example, user A can join with different users that craftily meet in a café to manufacture a specially appointed cell telephone system. Data sharing and correspondence can be further directed among the individuals from this "artful, physical-vicinity activated" group. At the point when A leaves the Espresso shop, the data he/she acquired from this entrepreneurial group (e.g., there will be an outside show in the Central Park tomorrow night) can be further dispersed to other newly formed crafty groups (e.g., with his/her partners in the working spot, or with different travelers on the transport he/she will take after). Same as friend-circle based exchanging, the infrastructure based astute exchanging is established on the scattering and coordinating of exchanging solicitations in sharp IoT situations. For instance, Bob needs to transfer a movie to Alice by means of the pioneering exchanging operators (OTA) running on his cellular telephone. While Bob visits every day with his friends, his exchanging solicitation is shared by friends circle (framing a pioneering group utilizing

978-1-5090-4065-0/16 $31.00 © 2016 IEEE
DOI 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.163

IEEE computer society

cell telephones). Since the moving extent and versatility example of Bob is generally settled (the quantity of individuals he can experience is therefore restricted), to expand the quantity of exchanging solicitation collectors and rate up the solicitation scattering procedure, OTA will utilize other portable hubs as "Friends" forward Bob's solicitation. After a day Alice who was out of town might be found by OTA.

In this paper, we will introduce the astute IoT, which addresses data dispersion and sharing inside and among entrepreneurial groups (with sets of gadgets) that are framed in the light of the development and pioneering contact nature of human. We delineate the idea of shrewd IoT through the accompanying "astute exchanging" user case. Especially we have exploited on reducing energy consumption by using the opportunistic IoT devices' movability for tethering communications when they encounter, substituting the traditional communications through the infrastructure for those delay tolerant services.

The rest of the paper are organized as below. The section II have reviewed the recent advancements in the areas of IoT, delay tolerant network, and mobility models and their impact on energy consumption which are fostering the CatchMe paradigm. Then we present CatchMe by proposing a time-space matrix to capture human mobility preferences as well as introduce the similarity analytical framework for mobile encountering analysis in section III. We have conducted three case studies in section IV to reveal that CatchMe is feasible and has great promises on substituting infrastructure-based transmission approach for those delay tolerant data services. Finally, the main contributions and also the layout of future work are drown in section V.

II. RELATED WORK

The mobile communications services are principally depending on the remote infrastructure support such as cell systems and WLAN [5]. The opportunistic IoT, nonetheless, addresses the restriction of remote infrastructures, for example, lacking system coverage opportunity and high cost of energy. The opportunistic IoT can take pervasive processing further, to investigate the human conduct and social association with upgrade opportunistic information sharing. The administration of opportunistic systems is based on the unconstrained availability between clients with remote gadgets [6], encouraging information directing, sending and exchanging between gadgets. It expands the opportunistic systems administration idea from two perspectives: (i) it is established from the Internet of Things vision, which acquires the way of smart things on encompassing detecting. Thus the local sensed data (e.g., activity motion, communications levels) can in this way be opportunistically shared to others, i.e., supporting the alleged participatory detecting, and (ii) it especially investigates the existing get-together phenomena of opportunistic groups [7] in the physical world and human societies, as well as studies the association and coordinated efforts among heterogeneous human groups.

The recent investigations of human movement have demonstrated that human movement conduct can be portrayed by utilizing truncated force law flights. In [8], the researchers talked about the mapping of human directions to the power-law flights, called Levy flights. On a basic level, the perception demonstrated that individuals tend to travel a large portion of the time over short separations, though a few individuals travel once in a while over many kilometers. The self-comparable Slightest Action Walk (SLAW) mobility model depends on comparative measurable examples joined with a minimum activity trip arranging calculation that depends on the perception that individuals endeavor to minimize the aggregate voyaging separation by going by adjacent destinations first [9]. The briefest ways on the city map to the following destination relies on upon the chosen transport methodology, for example, strolling or passing via auto or by transport, which characterizes, for instance, the velocity. Gotten from urban activity arranging research, a meaning of a mobility model considering parts of urban mobility are determined in [10]. Here, the mobility of people on foot within a workday is displayed in point of interest by utilizing an operator's model depicting motion between moving nodes and parameters portraying exercises. Since this methodology portrays additionally the connection between mobility attributes and exercises, it contributes to the field of movement demonstrating on the premise of genuine mobility information. In [11], the affectability of connectivity measurements to changes of mobility parameters has been researched.

In the study [12], a hyper graph based methodology is proposed having vertices and edges. It catches the high request cooperation in communicational and online networking having pair wise relationships. The most brief path issue is here spoken to by a hyper graph of exceptional qualities and element in nature. It can have the capacity to plot the topologies change likewise in the graphs which makes a difference in right path estimation based on their weight values. A portion of the researchers concentrated on the influence of most brief path distinguishing proof in diverse sorts of impromptu systems. The researcher in [13] have added to the protocol called Energy Based Routing Algorithm (EBRA). They coordinated the Dynamic Source Routing (DSR) convention to guarantee the base vitality consumption rate. The proposed mechanism comprises of three stages: nodes vitality consumption is restricted with the high portability; the impact of noxious conduct is lessened to maintain a strategic distance from the replaying of parcels and the unauthenticated node is recognized utilizing the advanced mark check. The reform results demonstrate that the proposed mechanism accomplishes less energy consumption rate, more strength efficiency, better throughput, less overhead and delay in the vicinity of the malicious nodes than the current. The studies in [14-15] take hereditary calculations to tackle the issue. In this variable-length chromosomes furthermore, their qualities have been utilized for encoding the issue. The study in [16] proposed a methodology actualized by presenting an edge esteem on every hub and transmitting the equivalent length of parcel on the course. The reenactment results exhibited in [17] confirm the viability of the proposed approach. The researchers in [18] have distinguished the bundles in charge of expanding energy consumption with steering protocols utilizing distinctive activity models. A correlation of the energy consumption of different protocols under CBR activity was the subject of work. The study [19] has looked at the energy consumption of two reactive protocols (AODV and DSR) under Pareto and Exponential movement. All energy devoured by every hub in transmission and gathering procedure has been assessed as the capacity of respite time, pace, number of nodes and sources, sending rate and zone shape. An investigation of these studies demonstrates that their regular objective is to enhance the energy consumption of directing protocols. On the other hand, the parameters mixed over by each of them is distinctive.

Additionally, the device to device (D2D) communications technologies have been developed recently, and the smart devices in close proximity can interface with each other directly and structure a communication network. Data traffic among the devices can be offloaded to the D2D network instead of transmission through infrastructures such as base stations. For instance, by authorizing D2D communications, some clients download substances from BSs while others might recover substances from their associates. Along these lines, D2D communications significantly reduces the limit requests in BSs and likewise reduces BS energy consumption [20]. Future wireless access networks will coordinate various radio access advancements (RAT), e.g., LTE and D2D near field communications. Subsequently, users will have various decisions on selecting RAT to augment their utilities. It is pivotal to adjust traffic loads among RATs to completely explore the limit of the heterogeneous wireless access networks [21] and also to minimize their energy consumption.

III. CatchMe and Mobility Similarity Analysis

A. Energy consumption model

A network can be represented by a graph $G\ (N,L)$, where N is the number of nodes and L is all direct links $(i,j),\ i,j \in N$. For prototyping the key ideas, we assume a general and simplified energy consumption model for the wireless or wired energy dissipation where the transmitter dissipates power to generate the radio or line electronics, the power amplifier consume energy to transmit the traffic, and the receiver dissipates energy to receive and process the radio or line electronics, as shown in Fig.1 below.

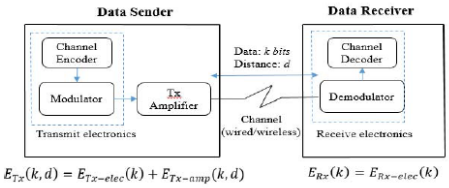


Fig. 1 A general energy consumption model of communications

Taking radio transmission as an instance, the power control can be used to remedy the signal propagation loss by appropriately setting the power amplifier. For example, if the transmission distance is less than a thresholdd_0, the free space prorogation model with the attenuation parameter of ε_{fx} is used, otherwise the multipath (mp) propagation model with the attenuation parameter of ε_{mp} is used. For the sender to transmit a volume of k-bits data to the receiver where is a distance of d away, the energy consumption model can be calculated as below:

$$E_{Tx}(k,d) = E_{Tx-elec}(k) + E_{Tx-amp}(k,d) \qquad (1)$$

For the case of radio transmission,

$$E_{Tx}(k,d) = \begin{cases} k \cdot E_{elec} + k \cdot \varepsilon_{fx} \cdot d^2, d < d_0 \\ k \cdot E_{elec} + k \cdot \varepsilon_{mp} \cdot d^4, d > d_0 \end{cases} \qquad (2)$$

where E_{elec} is the energy consumed by the transmitter and it depends on the factors such as the digital coding, modulation and filtering signal processing procedures. As for the energy consumed by the amplifier, it depends on the distance to the receiver and the acceptable bit-error rate. In addition, for the data receiver, the energy consumption can be calculated as:

$$E_{Rx}(k) = E_{Rx-elec}(k) = k \cdot E_{elec} \qquad (3)$$

Based on the above general energy consumption model for communications, we can see that the volume of data, i.e., k and the transmission distance of d are two critical but might be changeable factors which can vary the overall energy consumption, comparing to the energy consumed by the electronic components and also signal processing mechanism in the transmitter, receiver and also the relay amplifiers. *Is it possible to decrease the transmission distance d thus reduce the energy consumed by the transmission process*? In other words, the more transmission distance can be shorted, the more energy consumption can be reduced. This is the motivation triggering us to propose the Context-aware tethering communications hangout using Mobile encounter (CatchMe) mechanism, by fully utilizing the mobility of human (i.e., mobile users) and thus associated IoT devices they carried to implement data transmissions. Especially for those communications services, e.g., file transfer which has delay tolerant natures.

B. CatchMe paradigm

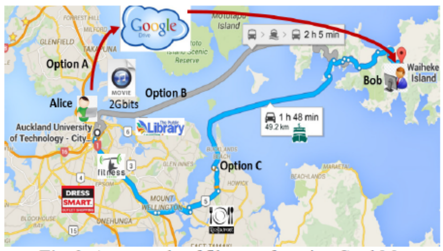


Fig. 2. An example of file transfer using CatchMe

In the previous session we have identified that the data transmission distance is a key but can be changeable factor contributing to the overall energy consumption. It is sounding that the longer the transmission distance, the higher of the total energy consumption. For example, we have two users Alice and Bob: Alice is working in a university located in the city, and she live there too. Bob is a freelance writer who is living in the Waiheke Island where is 50km far away from the city, and most of time Bob is working at home. One day, Bob is planning to write a story using Alice's case and Alice needs to send a video (e.g., 2Gbits) to Bob for reference. While Bob does not need the video urgently because he is working on something else and can only start looking at the video one week later. In this case, how can Alice send this 2Gbits video to Bob? There are several options available:

- **Option A**, Alice uses the cloud based storage service such as Dropbox, which is the most popular way we use every day: to upload the file to Dropbox, then share the link with Bob, who can use the link to download video;
- **Option B**, if direct file transfer service is available, for example, they both has own routers so can exchange data to each other directly;

The above two options are the most popular ways which are taken for granted in daily use, and their overall energy consumptions are mainly depend on the distance of the data transmission and also the size of the file.

Is there a 3rd option if considering the delay tolerant requirement (i.e., within a week of 7 days) of receiving this *video by Bob?* Occasionally, Alice and Bob have a phone call to discuss the case, and later on start chatting their daily life and hobbies. The conversations disclose their weekly mobility preferences such as going to the same Gym and library, as well as loving the same restaurant and shopping places. Even more, Alice releases that she has a scheduled trip with family to Waiheke Island during the coming weekend. Now they are all just aware that, actually there are many opportunities existing to meet each other during the coming week or say mobile encountering in several locations they both will visit so get a chance to pass the video file directly. This is the motivation triggering the proposal of Context-aware tethering communications hangout using Mobile encounter (CatchMe) approach. By knowing additional mobility information between peers, CatchMe can create their mobile encounter opportunities for exchanging information and the energy saving is to be significant (see the numerical study at end). Alice and Bob can have more options to exchange video file:

- **Option C**: Alice has a scheduled trip to Waiheke Island, so she can add an extra task to pass the file to Bob directly;
- **Option X**: By sharing more about their mobility plans, Alice and Bob can have several opportunities to meet each other to pass the video file

Rather than using infrastructure based transmission approach, after knowing the mobility of each other, Alice can use options C or X, by using their associated devices to carry the data file and then use their mobile encounter opportunities to transfer the video file to Bob but still satisfying the time requirement (i.e., within one week).

C. Time-space matrix and mobility similarity analysis

j: locations

Alice (*i*: time period)

	L1	L2	L3	L4	L5	L6
Mon	0.8	0	0	0	0	0.2
Tue	0.8	0	0.2	0	0	0
Wed	0.8	0	0	0	0	0.2
Thu	0.8	0	0.2	0	0	0
Fri	0.8	0	0	0	0	0.2
Sat	0.4	0.2	0.2	0	0.2	0
Sun	0.6	0	0	0.4	0	0

$M_{i,j}$

i represents time period
j represents locations
$M_{i,j}$ represents the normalized time duration (over 12 hour e.g., 8am -8pm) for a user in time *i* at location *j*

Bob

	L1	L2	L3	L4	L5	L6
Mon	0	1	0	0	0	0
Tue	0.4	0.6	0	0	0	0
Wed	0	1	0	0	0	0
Thu	0	1	0	0	0	0
Fri	0.4	0.3	0	0	0	0.3
Sat	0	0.3	0.4	0	0.3	0
Sun	0	0.6	0	0.4	0	0

L1: AUT campus
L2: Waiheke island
L3: City fitness centre
L4: DressSmart Outlet
L5: Restaurant
L6: City library

Time-space matrix *M*

Fig. 3 Time-space matrix for mobile encounter

For better formulating above problems, we capture Alice and Bob's mobility behavior and preferences by proposing a time-space matrix *M* as shown in Fig.3 above. In this time-space matrix, each row vector describes the normalized percentage of time the user spends at each location on a measured time period e.g., from 8am to 8pm in a day, reflecting the most significant locations to the user who prefers to visit for a while. Depending on the cases, we can define the granularity of the time duration as an hour, a day or a week, as well as locations as a specific room, a building or a region of which the matrix provides flexibility and scalability to represent the problems. Moreover, the singular value decomposition (SVD) technique [22] can be deployed on a given user's time-space matrix *M*, so to identify the main mobility trend or say mobility preferences as below:

$$M = U \cdot \Sigma \cdot V^T \quad (4)$$

where a set of eigenvectors, $v_1, v_2, \ldots, v_{rank(V)}$ represents the important trends in the original matrix *M*, which can be obtained from matrix *V* associated with corresponding weights $w_{v1}, w_{v2}, \ldots w_{vrank(V)}$ calculated from the eigenvalues in matrix Σ. This set of vectors are referred as the mobility trend of the particular user of M_A, as they represent the important trends in user *A*'s mobility preferences.

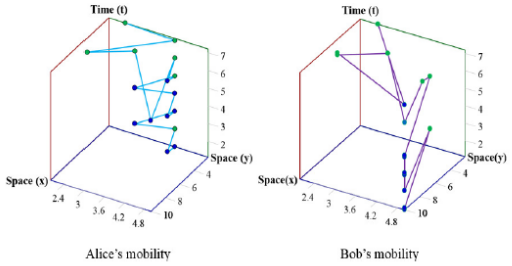


Fig. 4 Alice vs. Bob's mobility preferences (the points colored in green are the opportunities for their mobile encounter)

Furthermore, we propose a *mobile encounter index* metric, i.e., M_{ei} to indicate the similarity ranking (i.e., 0 represents totally dissimilar and 1 indicates exactly same) between two users' mobility preferences by quantitatively measure the similarity between the eigenvectors of their mobility matrices. For the pair of users, with respective eigenvectors as $A = \{a_1, a_2, a_3, \ldots a_{ra}\}$ and $B = \{b_1, b_2, b_3, \ldots b_{rb}\}$, the mobility similarity can be calculated by the weighted sum of pair wise inner product of their eigenvectors as:

$$M_{ei}\ (A,B) = S_{A,B} = \sum_{i=1}^{rank(A)} \sum_{j=1}^{rank(B)} w_{a_i} w_{b_j} |a_i b_j| \quad (5)$$

$M_{ei}(A,B)$ is a quantitative measure index that shows the closeness of two mobile peers in their mobility preferences of time-space dimension. The value of the similarity is within the range of $0 \leq S_{A,B} \leq 1$. The lower the similarity rank indicates that peers have very different time-space mobility preferences, in other words, they have very less opportunities to meet each other somewhere. Otherwise for the higher similarity rank cases, the user pairs have very similar mobility preferences, in other words, they have high probabilities to meet each other in some locations where they have same interests to visit even they do not aware of others might be around.

In addition, we can also quantify how the mobility preferences similarity between the same pair of mobile users varies with time. Firstly, we can calculate the similarity between two mobile users such as Alice and Bob, at two time points of $M_{ei}\ (A,B)_{T1}$ and $M_{ei}\ (A,B)_{T2}$, where *T* represents a specific time instance. We conduct this to all peer- pairs and then compute the mobility correlation coefficient of the similarity matrices obtained after a *T* interval as

$$M_{CC} = \frac{\sum_{A,B} (X - \bar{X})(Y - \bar{Y})}{N S_X S_Y} \quad (6)$$

where $X = S_{A,B}(T_1)$ and $Y = S_{A,B}(T_2)$ and the notations $\bar{X}$ and S_X denote the mean and standard deviation of *X*, respectively. *N* is the total number of peer pairs. This mobility correlation coefficient indicates how stable of the interactions between the peer pairs. It has been reported that the similarity metrics between pairs correlate reasonably well when the time sampling duration measured are not far apart. For the similarity analysis, we can further categorize the similarity measure as two types as with or without simultaneous temporal considerations then accommodating with different CatchMe communications strategies. Without simultaneous temporal case, it is not strictly requiring that peers must appear in the same location in the same time period. In this case, one user can temporally deposit the data into a storage box in that location, then when user B pass by can get these data which quite likes the way of physical mailbox works. It might require more energy consumption because it needs a relay point and extra storage resources. On the other hand, with simultaneous spatial-temporal case, it means two mobile users will meet in the same location at the same time which is perfect by mobile encountering. With above motivations and techniques, CatchMe has been proposed and it aims to fully utilize these mobile encounter opportunities to enable an infrastructure-free D2D communications by using IoT device's mobility. It can substitute the traditional data transmission through the network infrastructures so can save the energy.

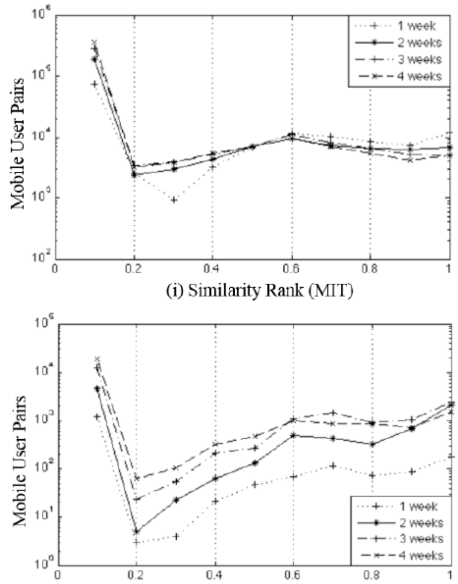
IV. Case Studies

In this section, we have conducted two case studies and one numerical study to proof-of-concept on CatchMe. We have calculated the mobility similarity distribution among mobile peers so as to investigate whether the mobile encounter phenomena is existing or not among mobile users. It is expected to see that a natural physical distribution of mobile peers who uses mobile devices turning into reasonably connected time-space correlated clusters. Each of these clusters has the mobile peers with similar time-space mobility preferences or say visiting the similar locations with a reasonable frequence. We have examined the real mobility traces collected from two university campuses which are publicly available at the CRAWDAD repository [23]. The information about these datasets is listed in Table 1 below.

Table 1: Mobile traces dataset

Campus	# Users	Duration
MIT	1350	Fall 2006
Dartmouth	1500	Fall 2007

These mobility traces are extensive with high density of active users and also include location information. Firstly, we have abstracted out a subset of peers from the population for further processing. Then we extracted the relevant statistics of these mobile peers' time-space mobility data including time and space location information. In the third step, for each mobile peer we have formulated the normalized time-space matrix as shown in Fig 3 with time granularity of one day. Then we deploy SVD technique on these matrices to extract their dominant trends. Finally, the mobility similarity analysis on any user-pairs has been conducted based on the duration of 1, 2, 3, and 4 weeks to validate the stability of our findings.



(i) Similarity Rank (MIT)

(ii) Similarity Rank (Dartmouth)

Fig.5 Similarity analysis of two mobility traces

The similarity ranking distribution histograms for those two university datasets are shown in Fig.5 above. It can be seen that, the number of peers as a function of similarity rank which quantifies their mobility similarity. It shows that, i) the mobile user clusters are existing and forming by peers with various time-space mobility similarities; ii) for those four time periods i.e., 1 - 4 weeks, a consistency and stability can be identified for the similarity ranking among mobile peers. The lower similarity rankings e.g., (0.0- 0.1), indicates a substantial number of mobile users is time-space mobility dissimilar or say they do not have similar preference to visit locations during the week. While on the other hand, high similarity ranks e.g., (0.9 – 1.0) are existing for some mobile peers, indicating the significant likelihood of close relationships has created tight clusters or say they have very similar mobility preferences to visit some locations more often. The fluctuations in the middle range of similarity ranking interprets partially similar and partially dissimilar peers. All of these are significant and provide an insight into the presence of mobile user clusters or say communities among peers, thus the mobile encounter opportunities are highly available in the network with similar location visiting preferences. These provide great potential of using CatchMe approach, especially for those ICT services with a nature of delay tolerant.

Moreover, we conducted a numerical study by using the video file transfer example described in Fig.2. The energy consumption for those four file transfer options are calculated as shown in Table 2 below:

Table 2: Energy consumption for four file transfer options

Options	Distance (km)	Communications Approach	Energy Intensity	Total Energy Consumption (*J*)	Reference
A	10500 (assume Dropbox servers located in US)	Cloud Storage and line tranmission	0.2kWh/GB*	$2.8x10^6$	*[24]
B	50	Line transmission	0.2kWh/GB	$4.57x10^3$	
C	<0.01	wireless D2D e.g. Bluetooth	$2.5x10^{-7}$J/bits*	0.5	*[25]
X	<0.01	opportunistic D2D e.g. Bluetooth	$2.5x10^{-7}$J/bits	0.5	

The numbers of energy intensity for traditional infrastructure based transmission and D2D e.g., Bluetooth transmission are referenced in [24] and [25] respectively. We can see that the results for different options varies significantly, in the 1-6 power orders of their magnitudes. The infrastructure-based option A, which is popularly used in a daily base, actually has the most heavy energy and resource consumptions i.e., $2.8x10^6$ J, and option B (i.e., direct file transfer) consumes energy much less i.e., $4.57x10^3$ J because shortening the communications distance. After knowing the mobility preferences information among the users, the CatchMe approach (i.e., Option C or X) can be possible deployed to achieve the same goal i.e., Alice sends a 2Gbit video file to Bob within 7 days, but it significantly reduce the energy consumptions (i.e., 0.5 J), 6 orders less to option A.

Here we need to notice that, the above case is only a simply numerical benefit demonstration and more complex protocols such as context-aware, automatic routing inquires embedded in IoT devices should be further developed to enable CatchMe mechanism operating automatically and intelligently. Even later on the security and privacy issues are to be addressed because the sensitive disclosure of mutual-mobility preferences are needed for CatchMe to make more sustainable selection decision among different communications options. The current state of our research is more proof-of-concept on CatchMe paradigm does not yet lend itself to sweeping prescription, but this novel way of CatchMe communications is thought-provoking and opening a new conversation for researchers to rethink and redesign more sustainable communications by fully utilizing the IoT mobility.

V. Conclusion and Future Work

A novel CatchMe approach is presented, in which a mobile encounter between the communicating pairs is sought to be established to directly exchanging those information with delay tolerant nature. Rather than using the infrastructure based transmission approach, CatchMe aims to utilize the movability of humans and/or the IoT devices to reduce the transmission distance thus its associated resources such as energy as much as possible. A flexible and scalable time-space matrix has been proposed to capture the mobility preferences of each mobile peer and the similarity analysis framework are introduced to reveal the mobile encounter opportunities between two possible communicating peers. Based on the mobility traces documented from two campus, two similarity case studies have been conducted and the results have confirmed that there are reasonable mobile encounter opportunities existing among mobile peers and CatchMe is feasible. More extensive work to develop intelligent communications protocols and algorithms for enabling automatic mobile encountering among peers are underway. The more complex mobility models and also the unpredictability of things' movements sometime should be considered to develop more adaptable CatchMe algorithms.

References

[1] Zhong N, Ma JH, Huang RH, Liu JM, Yao YY, Zhang YX, et al. Research challenges and perspectives on wisdom web of things (W2T). Journal of Supercomputing.

[2] Atzoria L, Iera A, Morabito G. The internet of things: a survey. Computer Networks 2010;54(15):2787–805.

[3] Boldrini C, Conti M, Delmastro F, Passarella A. Context- and social-aware middleware for opportunistic networks. Journal of Network and Computer Applications 2010;33(5):525–41.

[4] Hummel KA, Hess A. Estimating human movement activities for opportunistic networking: a study of movement features, In WOWMOM 2011: 2011 IEEE Int. Symp. on a World of Wireless, Mobile and Multimedia Networks, Lucca, Italy, 2011; 1–7.

[5] Fischer G. User modeling in human computer interaction. User Modeling and User-Adapted Interaction 2001;11:65–86.

[6] Boldrini C, Conti M, Delmastro F, Passarella A. Context- and social-aware middleware for opportunistic networks. Journal of Network and Computer Applications 2010;33(5):525–41.

[7] Campbell AT, Dartmouth C, Hanover NH, Eisenman SB, Lane ND, Miluzzo E, et al. The rise of people-centric sensing. IEEE Internet Computing 2008;12(4): 12–21.

[8] Gonzalez MC, Hidalgo CA, Barabasi A-L. Understanding individual human mobility patterns.Nature 2008; 453(7196): 779–782.

[9] Lee K, Hong S, Kim SJ, Rhee I, Chong S. SLAW:a mobility model for human walks, In INFOCOM '09: IEEE Int. Conf. on Computer Communications, Rio de Janeiro, Brazil, 2009; 855–863.

[10] Kim J, Sridhara V, Bohacek S. Realistic mobility simulation of urban mesh networks. Ad Hoc Networks 2009; 7(2): 411–430.

[11] Helgason O, Kouyoumdjieva ST, Karlsson G. Does mobility matter. In WONS '10: 7th International Conferenceon Wireless On-demand Network Systems and Services, 2010; 9–16.

[12] J Gao, Q Zhao, W Ren and A Swami, ―Dynamic Shortest Path Algorithms for Hyper graphsǁ, in Army Research Laboratory Network Science CTA under Cooperative Agreement W911NF- 09-2-0053

[13] S. Gopinath,.A.Rajaram, and N.Suresh Kumar, "Improving minimum energy consumption in ad hoc networks under different scenarios", International Journal of Advanced And Innovative Research (IJAIR) , pp. 40-46, September 2012.

[14] G.Singh, R.Gupta and A. Kushwaha, Implementation of Shortest Path in Packet Switching Network Using Genetic Algorithmǁ, in International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 2, February 2012.

[15] N A. S. Alwan, I K. Ibraheem &, S M. Shukr, ―Fast Computation of the Shortest Path Problem through Simultaneous Forward and Backward Systolic Dynamic Programming ― in International Journal of Computer Applications , Volume 54– No.1, September 2012.

[16] S. Suganya and S. Palaniammal, "A dynamic approach to optimize energy consumption in mobile ad hoc network", European Journal of Scientific Research Vol. 85 No 2, pp. 225- 232, September, 2012.

[17] W. Feng and J. Elmirghani, "Lifetime evaluation in energy-efficient rectangular ad hoc wireless networks," INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS, vol.23, pp. 1500-1520, 2010.

[18] S. Dubey and R., "Energy consumption using traffic models for MANET routing protocols", International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) Volume-1, Issue-1, pp. 84-89, 2011.

[19] Dhiraj Nitnaware and Ajay Verma, "Performance evaluation of energy consumption of reactive protocols under self-similar traffic", International Journal of Computer Science & Communication, Vol. 1, No. 1, pp. 67-71, January-June 2010.

[20] ONE simulator: https://github.com/robgjansen/tps-one-simulator

[21] Kim, M., Kotz, D. and Kim, S. Extracting mobility model from real user traces. IEEE INFOCOM, April, 2006.

[22] R. Horn and C. Johnson, Matrix Analysis, Cambridge University Press,published 1990.

[23] CRAWDAD: A Community Resource for Archiving Wireless Data At Dartmouth. http://crawdad.cs.dartmouth.edu

[24] D. Schien, V. C. Coroama, L. M. Hilty, and C. Preist, "The energy intensity of the Internet: edge and core networks," in ICT Innovations for Sustainability, Springer, 2015, pp. 157–170.

[25] I. Oppermann, L. Stoica, A. Rabbachin, Z. Shelby, and J. Haapola, "UWB wireless sensor networks: UWEN-a practical example," Communications Magazine, IEEE, vol. 42, no. 12, pp. S27–S32, 2004.

A New Energy Efficient Big Data Dissemination Approach Using the Opportunistic D2D Communications

Ambreen Memon(✉), William Liu, and Adnan Al-Anbuky

School of Engineering, Computer and Mathematical Sciences,
Auckland University of Technology, Auckland, New Zealand
ambreen.memon@aut.ac.nz

Abstract. The emerging cyber-physical paradigm endeavours to unite all the physical objects embedded with electronics, software, sensors, and network connectivity to allow more direct interactions and information sharing between the physical and cyber worlds. While these massively connected devices and their associated communications can exponentially increase the data generation, transmission, and processing which consume a huge amount of energy and finally end up with harming the environment seriously. In this paper, we propose a solution for energy efficient data dissemination by using the opportunistic device-to-device (D2D) communications. Each sender can decide either use network infrastructure or through encountering the end-users according to the quality of service (QoS) requirements of each data demand and also the mobility behaviors of the users. These decisions are based on the time and location- traces of daily mobility routines and related activities of users and their social relationship. The case study, based on the similarity analysis of the mobility traces, has confirmed the rich opportunities for encountering among people, thus the proposed approach has great promises to reduce the energy consumption of big data dissemination.

Keywords: Big data · Opportunistic routing
Delay tolerant network · Energy efficient data dissemination
Similarity analysis

1 Introduction

The increase in the use of mobile devices has changed the way that users share data and ultimately leading to mobile traffic expansion exponentially since 2011. As per current expectation, more than 24 ex-bytes of mobile traffic will be navigating operators' networks by 2019, with 72% of this traffic being created by the interactive media [1]. In addition, the growing number of mobile devices and their communications have increased significantly the energy consumption.

Delay tolerant networks (DTN) perform store-carry-forward routing to deliver the data in an end-to-end fashion, although a continuous end-to-end

P. H. J. Chong et al. (Eds.): SmartGIFT 2018, LNICST 245, pp. 155–164, 2018.
https://doi.org/10.1007/978-3-319-94965-9_16

communication path may never exist between sender and destination devices. The integration of the infrastructure-based networks and DTN has shown the benefits since it can boost routing performance and offload traffic from the congested infrastructure networks.

This paper targets to develop a novel eco-friendly and sustainable data transmission approach for offloading the data traffic from the infrastructure to the opportunistic- and social- based device-to-device (D2D) communications. It can be performed by exploring the existing movements and spatial closeness relation among devices. To complement the traditional infrastructure-based data transmission, the new idea is to optimally piggyback data on the moving physical devices for data dissemination to achieve the energy reduction as well as to ensure the QoS requirements. In such way, the proposed approach can fully utilize the users' historical mobility traces to predict the next location. If they are close enough and also satisfy the QoS requirements, the data could be directly transferred using D2D communication which consumes less energy. Otherwise, data could be transferred through the infrastructure-based network.

The rest of the paper is organized as below. Section 2 has reviewed the recent advancements in the areas of D2D communication, delay tolerant network, mobility models and their impact on energy consumption. Section 3 presents the energy consumption model for data dissemination. We introduce the new data dissemination approach by using similarity analysis in Sect. 4. We have conducted a case study in Sect. 5 to validate the new approach and confirm its great promises on substituting the infrastructure-based transmission approach for those delay tolerant data services. Finally, we conclude the main contributions and also layout the future work in Sect. 6.

2 Related Work

In D2D communications, the devices in proximity can interface with each other directly and construct a communication network. Data traffic can be offloaded to the D2D network instead of transmission through the infrastructure based network. For instance, by authorizing D2D communications, some users can download the substances from the cellular base station (BS) while others could get the substances from their associates. Therefore, the D2D communications can significantly reduce the traffic congestions and also energy consumption in networks [2]. In future wireless access networks, balancing the traffic load among base stations can be accomplished through adjusting the user's BS affiliations [3].

At Massachusetts Institute of Technology (MIT), predicting the user behavior is one of their research themes. Time duration is recorded by mobile phones when adjacent to cell tower IDs and Bluetooth devices. While Bluetooth devices demonstrate different behaviors based on how are the devices related to each other. It has been found that the presence of business students in a similar area is performing the same activities [4]. Bluetooth signals were constructed within individual's house to check the accuracy of data transmission into locations by cell towers [5]. Estimating next location of the objectives by using the dynamic

Bayesian Network can reach a successful rate of 93% to 99%. In prediction, the next cell is a sequence of location that they investigated in communication areas, need to improve those resources for reservation and QoS requirements [6]. Zeibart et al. [7] predict that driving to the destinations, given a partially travelled route by calculating the probabilities of different possible routes. Bauer and Deru [8] suggest the ways of predicting future destination along with previous histories. The work in [9] has used the Naive Bayesian classifier based model, which consists of the time slot of days, weekends and 1–8 h. Everyday time stamp is divided into multiple records, that consist of a list of Bluetooth MAC addresses and locations for mobility prediction.

3 Energy Consumption Model

A network can be represented by a graph G (N, L), where N is the number of nodes and L is the direct links (i, j) or edges between graph nodes. For prototyping the key ideas, We assume a general and simplified energy consumption model for wireless or wired energy dissipation where the transmitter dissipates power to generate the radio or line electronics. The power amplifier then consumes energy to transmit the traffic, and the receiver dissipates energy to receive and process the radio or line electronics, as shown in Fig. 1. Taking radio transmission as an instance, the power control can be used to remedy the signal propagation loss by appropriately setting the power amplifier. For example, if the transmission distance is less than a threshold value d_0, the free space prorogation model with the attenuation parameter of ε_{fx} is used, otherwise the multi-path (mp) propagation model with the attenuation parameter of ε_{mp} is used. For the sender to transmit a volume of k-bits data to the receiver where there is a distance of d away, the energy consumption model can be calculated as below:

$$E_{Tx}(k,d) = E_{Tx-elec}(k) + E_{Tx-amp}(k,d) \tag{1}$$

$$E_{Tx}(k,d) = k \cdot E_{elec} + k \cdot \varepsilon_{fx} \cdot d^2, d < d_0 \quad = k \cdot E_{elec} + k \cdot \varepsilon_{mp} \cdot d^4, d > d_0 \tag{2}$$

where, E_{elec} is the energy consumed by the transmitter and it depends on the factors such as digital coding, modulation and filtering signal processing procedures. As for the energy consumed by amplifier, it depends on the distance to the receiver and the acceptable bit-error rate. Energy consumption for the received data can be calculated by:

$$E_{Rx}(k) = E_{(Rx-elec)}(k) = k \cdot E_{elec} \tag{3}$$

Based on the above general energy consumption model for communications, we can see that the volume of data k and the transmission distance d are two critical and changeable factors which can vary the overall energy consumption, compared to the energy consumed by the electronic components and signal processing mechanism in the transmitter, receiver and also the relay amplifiers. It is possible to reduce the transmission distance d which is being traversed through the infrastructure. In other words, the more transmission distance can

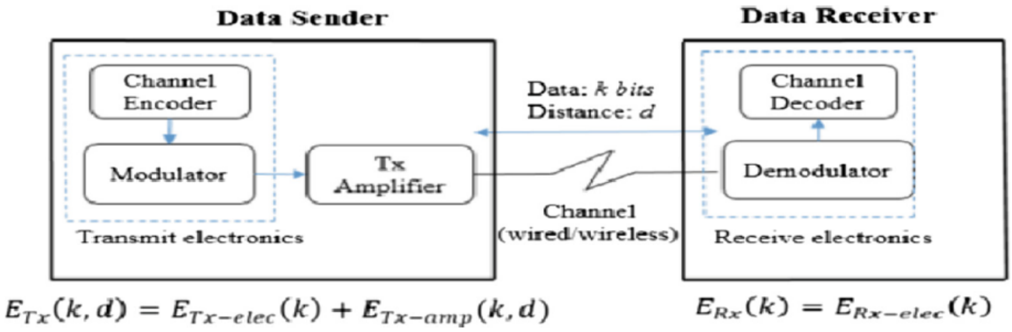


Fig. 1. Energy consumption model

be shortened, the more energy consumption can be reduced. This is the motivation triggering us to propose a new energy efficient data dissemination approach (EEDDA), by fully utilizing the mobility of human (i.e., mobile users) and D2D communications. They carried the data for delivery based on the prediction and users' mobility similarity analysis, especially for those communications services such as file transfer which has delay tolerant characteristic.

4 The Proposed EEDDA Approach

The proposed EEDDA approach, as shown in Fig. 2 has four processes including:

1. Data Collection
   In EEDDA approach, the first step is to collect the users' mobility data. In our case study, we reuse the data gathered by the project of Wireless Topology Discovery (WTD) that was handled at UCSD [10]. It has the traces of 300 people's accessibility of PDAs to WiFi. All the traces has two portions of discussion. One portion consisted of trace data. The other file contained the known locations with access points for local coordinates. Eleven-week trace duration started from the 22 September 2002 to 8 December 2002 was the data collection period.
2. Analysis
   The WTD has sampled and recorded the above information for all access points (APs) for every 40 s, which may fill in all its frequencies. The analysis of the collected data was conducted while running on a student's device. During a sample, the above information is given by WTD for all sensed samples. In Fig. 3 below, the three entries were recorded if a device has three APs in one sample (the entries include the IP address, signal strength, and attached flag). To extract basic records that show the user's location, the Associate field is used in this study. The user's device is located near the connected access point (Associated = 1) is based on the assumption. A list of neighboring access points is created for the sensitive access points that were not selected for the Association (Associated = 0). To record that a user should be decided at any time, AC-POWER field can be used for it. The assumption is that

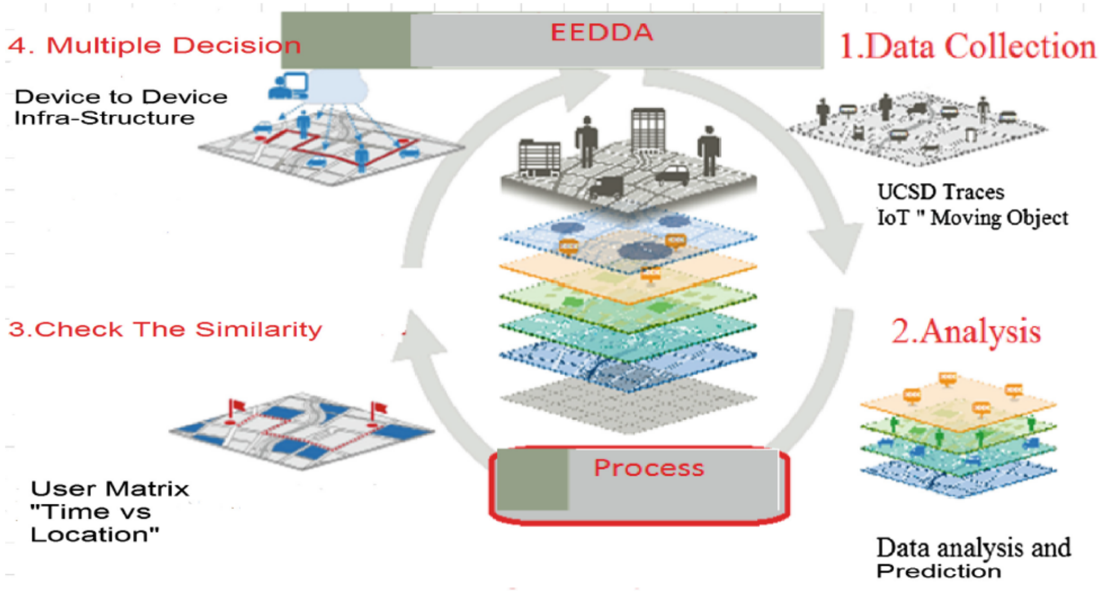


Fig. 2. The Proposed EEDDA approach

USER_ID	SAMPLE_DATE & TIME		AP_ID	SIG_STRENGTH	AC_POWER	ASSOCIATED
123	Sep-22	0:00:00	359	8	0	0
123	Sep-22	0:00:00	363	5	0	0
123	Sep-22	0:00:00	365	11	0	1

Fig. 3. The fields and usage of the database

the user is not mobile and AC-POWER = 1 is a plug-in in the device. In weight gain of individual access points, the given time, signaling power SIG-STRENGTH can be used.

In different location alignment algorithms, the SIG-STRENGTH field can be used. For example, a user is using the trilateration algorithm then it is between all access points. The AP-ID uses only one assigned number to label every access point in the field databases. Here we use these values as a location label, and when evaluating future locations, these values are developed by the model. In a real application, AP-ID field will be backed up on the map or on a map named on a useful location. The SAMPLE-TIME contains the date and time. The recorded state of every 20 s and end of 11 weeks period by their own devices. Throughout the pre-processing, the week fragments at the start and end of during the 11 weeks were rejected so as to provide 10 whole weeks' samples with User ID field for each record to the specific user. This algorithm does not use the user information, but simply uses the user-id field for partitioning the logs into individual user logs. The prediction can be done by individual user but not on the entire group.

3. Check Similarity

   The sequence prediction is to predict the next item in a sequence, which can be considered as a kind of rating. There are potential results for the alphabet used to create a layout. They are known in advance and predict the model in which the next item is in sequence. First signify theorizing symbols S_1, S_2, S_3, S_4, S_n. The n represents the number of symbols in the alphabet. When training sequences are described with the symbol of t:

$$Where\ X_i \in \sum X_1, X_2, X_3 ... X_t \tag{4}$$

   this calculation defines the conditional probability

$$Pr \{X_{t+1} x_{t+1} | X_t = x_t, x_{t-1} = x_{t-1}\} \tag{5}$$

   This calculation has been used in stationary Markov chain [11]. In our case, we are considering the stationary one, because the probabilities are not only depend on the same time, even sub-sequences repeat at the same time but with different location in sequences for each repeat or shift S and for all X_i.

$$\begin{aligned} Pr \{X_1, X_2 = x_2 ..... x_n X_1, X_2 = x_2 ..... x_n\} \\ Pr \{X_{1+m} = X_1, X_{2+M} = x_2 .., X_{n+m} = x_n\} \end{aligned} \tag{6}$$

   This process is called Markov model because the probability is likely to be on the variable. The number used is the variable for variants, L, model length, or order. The pre-variable sub-division is called history or context. If the contextual length of the context is set continuous, the model is called fixed length Markov chain [12]. The variable length Markov chain with length L, the context length used to vary the maximum number of L on the prediction of the Markov channel. The first order of Markov model makes the basis for prediction model here.

   Moreover, the raw data from the WTD experiment combines all logs, from all operators into one file. The entire logs of all the operators are associated with the document by the raw data from the WTD test. The first step of associating all logs from different users into one folder was to divide the raw information into various documentation for an individual user. Records usually here refer to as sensed, non-associated, access points. Records with the same date and time had the maximum amount of indication matrix. Whereas the rest records were rejected. Reports with similar user name, access point and having a starting time within one minute of the preceding record's starting time are termed as contiguous records. The output showed the length of each session collected over a day. The statistics recorded by a full-time active mobile device is regarded as the first movement data. The data representing users' significant locations revealing social interacting applications is called the destination data. In this work, significant locations are determined solely by a length of stay of at least 10 min [12]. The destination data and movement could be considered as movement location and significant location

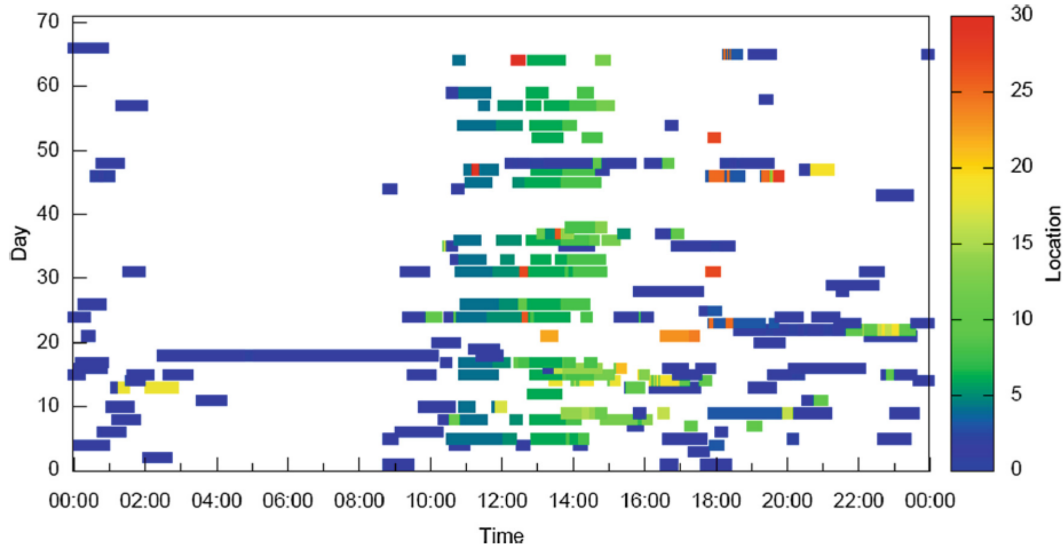


Fig. 4. Data (1-minute or less) for User 003 (Color figure online)

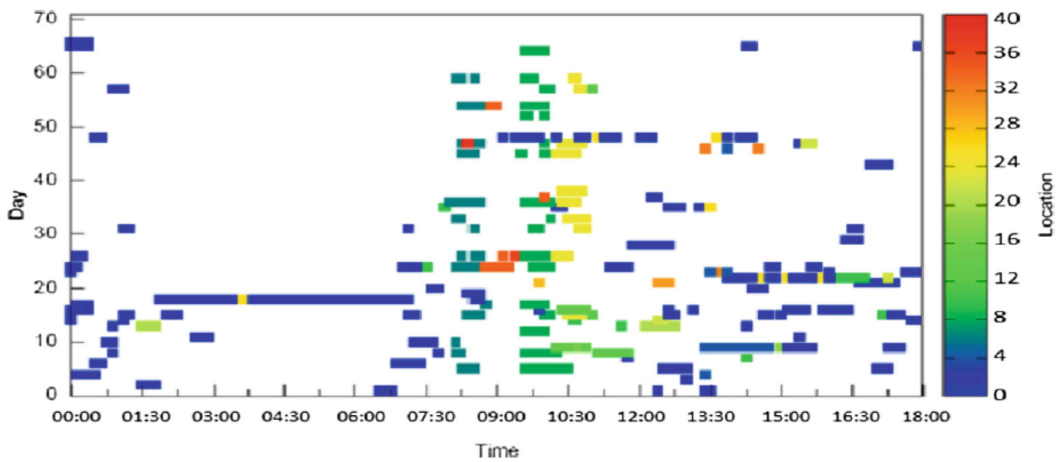


Fig. 5. Data (10-minute) for User 003

individually. The starting time is measured to be on entire number nearest to the last minute dataset of MoveLoc. Later on, all the duration include those, slightly lesser than 20 s, were included of one minute window. When the session started then it was considered as the highest period of one-minute time duration, declining the rest our work prediction on the future location and time. The Fig. 4 shows the MoveLoc data for user 3. The color bar shows that this user visited 40 locations, over the 10-weeks recording period. The x-axis represents the time of day. Anyone can observe that this user had some regular locations between 1:00 pm and 2:00 pm and some of the movement between locations is indicated by the change in colors/shades. Figure 5 depicts that User 3 MoveLoc (1-minute or less) and User 3 SigLoc (10- minute) data and dataset location. Where user consumed the minimum of 10 min, were covered in the SigLoc dataset. Elimination other sessions less than 10 min was the first phase of producing this assessment. The start-time of the other sessions were almost equalized to the bordering 10 min. The SigLoc data for User 3 is indicated in Fig. 4. One can notice that the number of locations fell from 40 to 20 and the transitions between locations were removed. Though the existing point of proof-of-concept on EEDDA, which does not advance towards a far-reaching position, yet the study outcome is relatively stimulating.

4. Multiple Decision
   This is the last process of EEDDA, of which the device will be able to take the decision as per the result of the similarity analysis. The data is either transmitted through the infrastructure based network or D2D communications. The Fig. 6 shows the flow chart of the decision process.

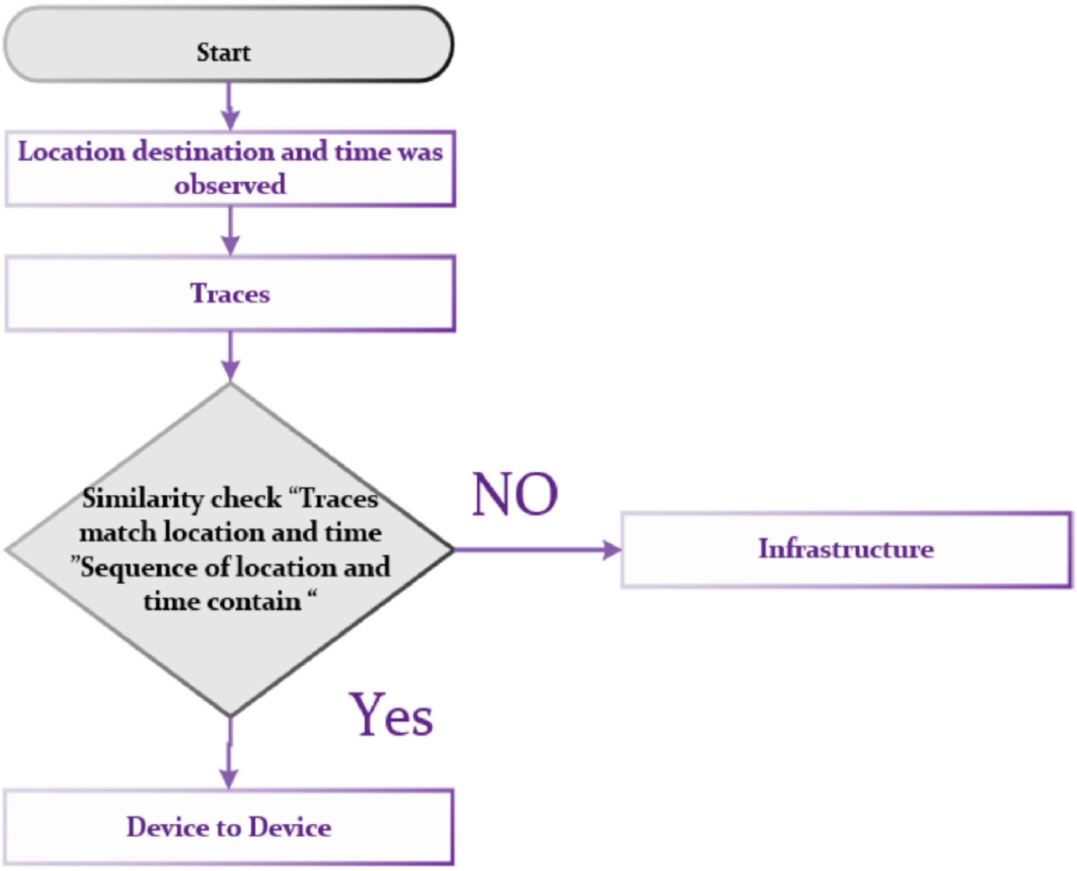


Fig. 6. The process of multiple decision

5 A Case Study

We have identified that the data transmission distance is a key but changeable factor contributing to the overall energy consumption. Here we consider a scenario, of which a professor in university and his students have communicated with each other very frequently. They are all in the same building but on the different floors. We assume that the data is delay tolerantable. In such way, we have two options including option A: send the data through infrastructure-based network, and option B: infrastructure-less approach i.e., D2D communications.

To disseminate data through these options, we assume further two possible ways: (a) Human mobility traces get matched according to the movement behavior. In such way, the proposed approach can fully utilize the benefits of mobility traces and check similarity of location and time. When they are matched, the data can transfer through D2D while consuming less energy. (b) In this option, we use the prediction model with human mobility traces that checks predicted location and time of user's movement. When it matches, the data will transfer by D2D communications.

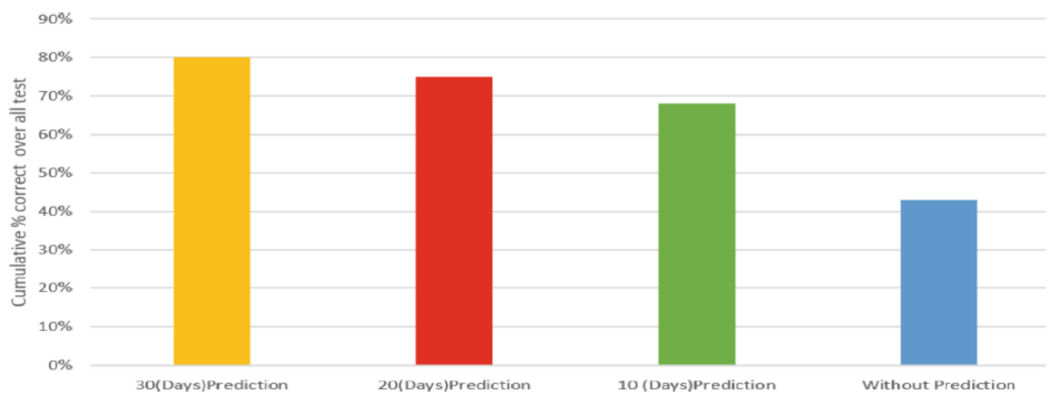


Fig. 7. Test result with 30 days, 20 days, 10 days interval prediction and without prediction

Figure 7 shows the correctness of the results as per time allocated for prediction. If we simply use the traces, the accuracy of the results is 45%, while considering delay for about 10 days interval. While in the prediction option, it gets increased by 70%. However, when we increase the interval of delay for about 20 days or 30 days, the accuracy could be better than the previous. This indicates that the longer tolerated delay and mobility duration, the higher opportunity that two users encounter each other to have D2D communications, which could reduce the energy consumption significantly.

6 Conclusion and Future Work

In this paper, a novel energy efficient data dissemination approach (EEDDA) is introduced, of which a mobile encounter between the communicating pairs is sought for directly exchanging data. The similarity analysis framework reveals the mobile encounter opportunities among communication peers. Peer device similarity depends on the moving ability and adaptability of interacting with the devices. The work is ongoing for developing communication protocol and algorithm but also for peer countering and automation. The greater number of complex mobility models and the unpredictability of devices' movements should be considered too to develop more adaptable EEDDA in future work.

References

1. Micinski, K.: Interaction-based security for mobile apps. Ph.D. dissertation, University of Maryland, College Park (2017)
2. Kjærgaard, M.B., Langdal, J., Godsk, T., Toftkjær, T.: Entracked: energy-efficient robust position tracking for mobile devices. In: Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, pp. 221–234. ACM (2009)
3. Dogar, F.R., Steenkiste, P., Papagiannaki, K.: Catnap: exploiting high bandwidth wireless interfaces to save energy for mobile devices. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp. 107–122. ACM (2010)
4. Ziebart, B.D., Maas, A.L., Dey, A.K., Bagnell, J.A.: Navigate like a cabbie: probabilistic reasoning from observed context-aware behavior. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 322–331. ACM (2008)
5. Elnekave, S., Last, M., Maimon, O.: Predicting future locations using clusters' centroids. In: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, p. 55. ACM (2007)
6. Roddick, J.F., Spiliopoulou, M.: A survey of temporal knowledge discovery paradigms and methods. IEEE Trans. Knowl. Data Eng. **14**(4), 750–767 (2002)
7. Mayrhofer, R., Radi, H., Ferscha, A.: Recognizing and predicting context by learning from user behavior. na (2003)
8. Holbach, H., Bourassa, M.A.: The effects of gap wind induced vorticity, the ITCZ, and monsoon trough on tropical cyclogenesis. In: 2012 International Ocean Vector Winds Science Team Meeting (2012)
9. Bellotti, V., Begole, B., Chi, E.H., Ducheneaut, N., Fang, J., Isaacs, E., King, T., Newman, M.W., Partridge, K., Price, B., et al.: Activity-based serendipitous recommendations with the magitti mobile leisure guide. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1157–1166. ACM (2008)
10. UCSD WTD Trace. http://sysnet.ucsd.edu/wtd/
11. Yang, G.: Discovering significant places from mobile phones – a mass market solution. In: Fuller, R., Koutsoukos, X.D. (eds.) MELT 2009. LNCS, vol. 5801, pp. 34–49. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04385-7_3
12. Vu, L., Do, Q., Nahrstedt, K.: Exploiting joint wifi/bluetooth trace to predict people movement (2010)

Paper 3

Sustainable Smart Connected Buildings: Random Forest based Energy Saving Model

Ambreen Memon[a,*], Jeff Kilby[a] and M.Ishtiaq[b]

[a]*School of Engineering, Computing and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand*
[b]*Department of Software Engineering, Foundation University Islamabad, Islamabad, Pakistan.*

ARTICLE INFO

Keywords:
similarity analysis
mobility
Device- to-Device communications
random forest model
Opportunistic connectivity

ABSTRACT

A smartphone may opportunistically pass data to another smartphone either as a direct destination or to use it as a carrier for passing the data to a third smartphone. This network of connectivity may be referred to as opportunistic network. Routing algorithms built for these networks aim to increase the probability of an effective transmission of messages. The most common method is to measure the probability of transmitting a message using details such as node contacts and information of area, while using past experiences to forecast potential ones. Conventional mechanisms for routing no longer suit. In this paper, we have proposed a random forest based model that used mobility traces of different users for future location prediction and then employed these future locations for encounter prediction. We have used university of Southern California (USC) mobility traces in the dataset for the experiments. The purpose of our proposed model is to reduce the resource consumption i.e. energy, and bandwidth during data transmission in daily movements. We have compared and analysed the Gaussian process and random forest by using real-world mobility data traces and show that how these encounter prediction results save 33% energy in comparison with other models. We calculate the device to device (D2D) and internet transmission cost and show that D2D is better than the internet.

1. Introduction

In recent years mobile phones and laptop PCs have gradually become commodity devices. They are seen everywhere, as more people understand for several purposes that possessing more sensing and processing capability is desirable in daily circumstances (1). Smart and connected communities as a research area are at the intersection of civil infrastructures in the social sciences and cyber-physical systems. The rapid and trans formative changes driven by innovations in smart sensors, which are now embedded in nearly every physical device and system we use, enable this research area. The results of such advances can be seen in a number of areas, including travel, energy, emergency management, healthcare, etc.(2).

A network composed of mobile devices used by individuals in understanding behavioral patterns can prove to be a key factor in optimizing the process. The people in their activity appear to obey trends. These trends can be seen in the interactions their mobile apps identify (3). Understanding whether to attack opportunistic networks and to whom a node will move a message to get met threshold and latency to the maximum extent. Predicting a device's potential experiences is thus of vital importance to enforce a successful opportunistic routing algorithm (4). In this paper, the energy consumption model is used to predict possible energy saving gains from building projects. The obtained results demonstrate that the proposed energy consumption model learns occupant behavioral patterns from the building. Our findings indicate that the proposed model of energy use knows about occupant behavioral trends from the building.

In fact, it reproduces them accurately, forecasts the strength of buildings energy usage and detection of possible carbon loss zones. We also propose the use of supervised learning techniques together with Random Forest Regressor and Gaussian process modelling to predict future encounters based on historical patterns of individual nodes. To equate the findings produced from the Gaussian and Random Forest model, we measured precision of predicted encounters at various locations by all methods. Such tests contrasted precision of predicted encounter in separate weeks of the month at a similar position. The highest accuracy achieved by Random forest model was for location 5 which was more than 80%. A Gaussian process is specified by a mean function or a covariance function (or the kernel). The random forest is a supervised learning algorithm which randomly creates and merges multiple decision trees into one forest. In this Research we propose a generalized method of predicting future encounters based on a machine language technique,

amemon@aut.ac.nz (A. Memon)
ORCID(s): 0000-0002-7302-3815 (A. Memon)

the Random forest classifier and compare with Gaussian process probabilistic classifier. For our experiments, we are using USC data traces. The traces are real measurements taken from the university campus of USC. The data contains information about Wi-Fi associations, user's profiles. The data is about six different buildings in the campus, access points, data about different days including weekdays and weekend.

The remainder of the paper is set out as follows; in section II we present the literature review. Section III presents the methodology. It explains the Random forest Model, the proposed Random forest for Encounter Prediction (RFMEP) also explained. It depicts the Gaussian Model, scenario for the Random Forest Model encounter prediction, and energy consumption model for sustainable energy approach. Section IV describes RFMEP work evaluation. It gives results for the encounter of different users through the Gaussian Random process and encounter prediction results, the comparison of the accuracy of the Gaussian and Random Forest Model is also described. Section V concludes the whole discussion and presents future work.

2. Literature Review

In the current revolutionized digital world, the number of smartphone users is increasing exponentially [4]. The proliferation of these devices is resulting in increased mobile traffic which has reached up to 79 % in 2019 (5). The mobile devices form a wireless mobile adhoc network (MANET) which provide different advantages of scalability and flexibility, as different users can join or leave the network without any prior notification (6). The users of these mobile devices travel at different places autonomously in any direction. This mobility of the mobile nodes enables them to dynamically and-+6+6 arbitrarily connect with each other. This mobility of users makes the MANET environment unpredictable for real-time applications of data transfer (7),(8). In(9), Prophet-Probabilistic Routing Protocol utilizing History of Encounters and Transitivity has been introduced in an article by (Lindgren, Doria, Schelen, 2003). The structure´ of this protocol for delivering a message is designed to use the transitivity and encountering history. It supposes that the nodes have a predictable repeated mode but not a random movement. Such a motion can result in the usage of delivering a quick and effective message.

Delivery predictability for an individually known destination can together be termed as a probabilistic metric. This probabilistic metric is calculated by each node before sending the message, and the probability of node A meeting node B is the probabilistic metric P(a,b)[0,1] Furthermore, predictability vectors are exchanged when two nodes meet each other. The carrier node only transfers the message to the neighbouring node when there is a high probability of the neighbouring node to meet the destination node. Otherwise, the message is not delivered further and kept to itself. Moreover, for buffer management at the nodes PRoPHET utilizes a FIFO queuing or communicating mechanism. After forwarding to a minor number of concentrated nodes, it can result in the consistent dropping of packets. It is independent of including any message scheme. Slow and inefficient responses to changing trends might be performed by this protocol in regards to nodes' movements. Additionally, for deciding the best carrier of message later on PRoPHET uses nodes' delivery predictability only. Whereas, for choosing the finest next hop to the destination, PRoPHET+(10) makes use of various parameters like Buffer, Location, Popularity, and Power. Certain weight assigns these values of parameters also, for obtaining an ultimate value from 0 to 1, these parameter values are normalized, By the obtained values of these parameters and their comparable weights, a final metric is calculated at this stage. The best next-hop is then determined by this final metric. Such particular protocol uses a lot of parameters for selecting the next hop. For these values, the calculations are hard which gives an outcome of different overhead in passing a message. Discussing further on, for predicting the upcoming behaviour in nodes' movement in the network, it does not use any parameter. In (11), Boldrini et al. Proposed HiBOp - 'history based routing protocol for opportunistic networks. To identify a better path, this protocol uses node s' current context. A screen-shot (snapshot) of the environment in which a node resides in currently. Identity Table (IT) and History Table (HT), both store the local-context. They are exchanged and current context (CC) is built when nodes meet one another. Previously, the value s that nodes see from the ITs are stored by the History Table. A Continuity Probability (Pc), Redundancy (R) and Heterogeneity (H) counters are related to every value. A node's every feasible context information is stored and used by HiBOp. These are very difficult to find generally. Withal, a huge measure of memory space on each node is required by IT, CC, RT, and H T. In (12) the topics and details of the discussion are similar to the description in (13).

In, Spyropoulos et al. proposed an extension to the Epidemic routing protocol by calling it Spray and Wait protocol. This protocol works by assisting to reduce the spread of the number of copies across the network. Hence, the congestion due to causes of flooding is reduced. In details, forwarding process has two phases which are called Spray Phase and

Wait Phase. When a message only spreads to L nodes, that is, L copies of messages are constructed or created, it is then called Spray Phase. Whereas, the relay nodes with a copy of message wait for the destination to directly transmit it. The chosen value of L indicates the probability of message delivery and is strongly responsible for this action. Again, the determined value of L is dependent on the network parameters. The protocol undergoes issues like delays and consumption of resources, despite the fact that L is selected to limit flooding. In (14) an improved version of spray-based routing schemes was presented by Makhlouta. This version of improvement can be classified by naming Adaptive Fuzzy Spray and Wait.

The aim of this study to predict pedestrian's similar location even researchers can revise the location update form a movement trajectory. Mobile phones or another device will sometimes be switched off, unremembered or entering urban canyons or other places with such low GPS coverage. User updates i.e. significant time lapses between locations updates are one of the crucial facts on those Applications such as Facebook(15)and Twitter are dependent. Additionally, to confirm maximum privacy policies users shall have access to record deletion or prevent recording as per choice (16). Bauer and Deru (17) Suggests two ways of predicting future destination along with previous histories. A Market Basket Survey of several purchasers can identify the connection between locations and produce rules. As for example, when someone goes to the coffee shop, he or she may buy the cake or biscuits this algorithm determines the same path the author on paper show, the algorithm but didn't show the appropriate results. Lymberopoulos, Bamis and Savvides (18),(19). At the cellular system of a based station, two algorithms namely PPM-C and LAST where used for anticipating the number system. A heuristic model that predicts two factors is considered to be LAST. One factor is no user movement and other is similarity of the former and next base stations. For a given portion of historical data, the correct algorithm to be applied is determined by the movement of Market Basket breakdown. This movement takes place through various pattern formation. For four-time steps, that the user remained connected to the similar base station is shown by historical data of Market Basket analysis. The LAST algorithm predicted the most recent base station accordingly. In cases, where the historical data showed the pattern of a base station, there is another case Market Basket Analysis did not find a common movement at that time. PPMC predictor was applied and those algorithms return the prediction on the single best location and this was author's hard decision. Author uses the vectors for all predictions refund, referred to as soft decision and aggregating the hard decision with the soft decision. Vu, Do, and Nahrsted at al results of their locations, period of time prediction recently published in (20). They used Nave Bayesian classifier based model, the model consists of time slot of days, weekends and time slots of 1 to 8 hours. Every day time stamp is divided into multiple records that consist of a list of Bluetooth MAC addresses and locations. When the MAC address match with people encountered by the user looking forward to the type and time of time slot predictions, predict the most likely (up to three) position, mean measurement, standard deviation and people (Bluetooth Max) mostly encountered at that location. They test the 50 users and achieved the results of 80 % using 2 hours' time slot. However, when they used the time slot of 1 to 8 hours, the result (21) was not significantly changed, which means the predictions of the most commonplace (for example, office) can be returned and places cannot be offered which have been briefed for short time. The collective movement patterns, daytime, land use and interest-point approach to a stable model (22).

It can predict 60 % right within an hour according to the next position of an individual. This work can be one of the first to add to the geographical models, and this geographical model movement pattern of the group to an individual for location predictions. Applied spatial temporal data mining to trajectories in order to predict the next area and arrival time of a vehicle. Monreale et al.(23) after encoding the change-overs between regions into a trajectory tree, the data of GPs location are put together according to the areas. The transition to child area from the parental area is denoted by each child. The child area supports in categorizing this transition. Moreover, it frames arrange of time for the movement to child area from parental area. For predicting upcoming areas, partial trajectories are matched in non-permanent tolerance with pathways.

In this way, the accuracy of prediction gets better than 54 %of models which can predict the time of arrival and accumulate vehicle behaviour to put on separate predictions. The algorithm in (24) uses duration at the last location to determine the next location, it requires continuous data use duration information and time of arrival to predict future locations.

3. Methodology

For reducing the Energy consumption, the current system first uses the mobility traces of a user having a mobile phone (node). These mobility patterns yields the time-space associations of the sender and receiver node. The time-space associations of both nodes are used for similarity ranking. Based on similarity profiles, the future location and encounter time interval is predicted for different users. In the last phase, the energy consumption model based on opportunistic Model works for reducing the energy consumption in the overall network.

Figure 1. Shows overall working how our RFEPM work predicts encounter among users and help to reduce the energy consumption.

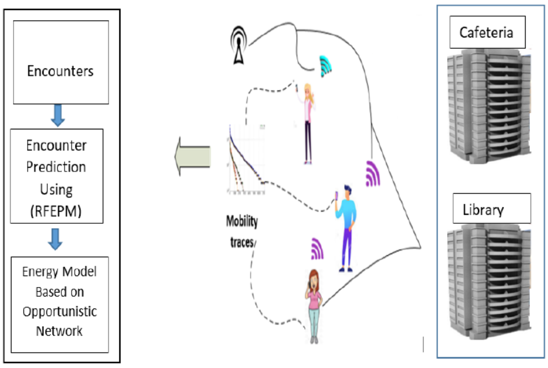


Figure 1: Model diagram.

Figure 2 shows the flow chart that explains the overall working of the RFEPM work of encounter prediction.

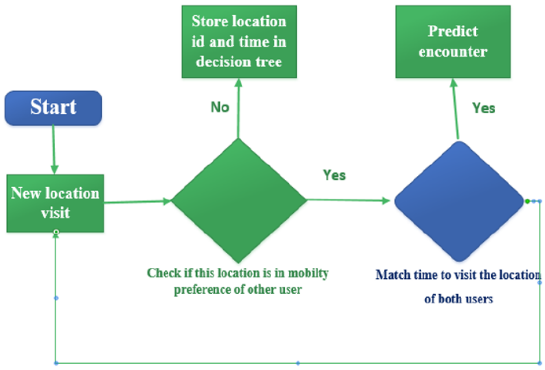


Figure 2: Flow chart explains the overall working of the proposed work of encounter prediction.

If a user visits a new location, a check is performed to see if this location is also in the mobility preference of some other user, if no other user is found with the mobility preference for this location, the location is added into the mobility history of the user. For adding the mobility history, the location id, and time for visiting the location is stored in the new sub decision tree. If, a user is found with the mobility history of location, the time for visiting the location is matched with the visiting time of the first user. In case of a match, an encounter among these users is predicted at the matching location and visiting time.

Before understanding the further working, there is a need to first understand what is the used Random forest model, how it works, and store the data.

3.1. Random forest Model

Random forest model is based on the concepts of decision trees, where each tree store particular information and grows in accordance with a specific parameter. For making the predictions, the data from ensemble is accumulated. Random forest trees produce quick output, are easy to implement, and are able to deal with numerous input variables without the problem of over fitting. In this model, each tree in the forest is created when at first, for each node, input features are selected and then further divided. The best split is calculated based on the input features in the training Dataset. The tree grows further in the same manner without pruning. The overall outcome is calculated by taking average of individual outcomes of the sub trees.

3.2. Random forest encounter prediction model (RFEPM) for energy consumption

Before formalizing the model, some definitions are being presented. First, select some parameters which are suitable for the dataset and these parameters help us to improve the accuracy of the random forest model. Throughout the document, suppose that the given training sample data input features $F_n =$ of the dataset with$[0,1]^e$•Rvalued random variables of encounter $(e \geq 2)$ with the same distribution as an independent generic pair (F_n). For fixed $u \in [0,1]^e$, the goal is to estimate the classification function $r(u) = E[F|u]$ using the feature F_n.

Tree Generation: -Formally, a random forest is a predictor consisting of a collection of randomized base decision trees $[140m \geq 1][$, where$E_1, E_2, ...$are i.i.d. outputs of a randomizing variable E. These random trees are combined to form the aggregated classification estimate.

$$r_n(u, F_n) = E_E[r_n(u, E, F_n] \tag{1}$$

Where in equation 1, E_E denotes expectation of an encounter concerning the random parameter, conditionally on users u and the features of dataset F_n. In the following, to lighten notation a little, the dependency of the estimates in the sample will be omitted, and write for example $r_n(U)$ instead of $r_n(u, F_n)$. The random forest model for encounter prediction consists of a set of random decision trees. Each $j^t h$ random treer^j has a co-ordinate U and variable E which determine the location for performing cuts for the new tree. Here in equation 2, n is the size of the sample and $_n(\varepsilon U, E\varepsilon)$ is a time whose occurrence resulted in the cuts for the new tree. Let$C_n(U, E)$ is the cell obtained after each random partition. Each random tree will be:

$$r_j = \frac{\sum_{j=1}^{n} O_j 1[\mu_j \in C_n(U, E)]}{\sum_{j=1}^{n} 1[\mu_j \in C_n(U, E)]} 1_{\mu n}(U, E) \tag{2}$$

Where O_j is the output of the tree. Assume, the data is spread in a Euclidean space, and each random tree explores a sub-section of that space. Each tree caters to the region of the search space in a discrete random manner. The selection of location sub-space is itself a random process. The selection of random, sub-space is justified by the notion that is combined with the effect of bagging will result in improved performance. In the equation 3, the performance of the random tree depends on the selection of sub-space and the cut of the tree applied over successive steps. The time $\mu_n(U, E)$ is

$$\mu_n(U, E) = [\sum_{(} j = 1)^n 1_{[} \mu_j \in C_n(U, E)] \tag{3}$$

In equation 4, After considring the final expectation E_f the estimate for each random tree will be

$$r_j(U) = E_f[r_j(U, E) = E_f \frac{\sum_{j=1}^{n} O_j 1[\mu_j \in C_n(U, E)]}{\sum_{j=1}^{n} 1[\mu_j \in C_n(U, E)]} 1_{\mu n}(U, E) \tag{4}$$

in the next step, the correlation among different trees is calculated. The correlation of different trees is calculated using equation 4.

$$W =_(i = 1)^P [R_i(U + \in)/P] \tag{5}$$

Where W is the variable of interest that is to be predicted. P is the no. of estimators used in the search process, R_i is the classification rule been set for predicting the value of W. U_i belongs to set U, which is a learning set. $\in$ is a zero-mean random noise in the data. The definition of $\in$ is given in equation 6

$$\in = [\beta[\in |U] = 0] and \beta[\in^2 |U] \tag{6}$$

is finite

3.3. Gaussian process for encounter prediction model

The existing work for encounter prediction use Gaussian model. A Gaussian process is a collection of random variables. Any finite set consisting of these variables possess a joint Gaussian distribution. This model can also be defined as a distribution over functions. The classification made by Gaussian process is non-parametric. This method is mostly used for opportunistic mobile networks (OMNs), where connections are sparse. A Gaussian process is

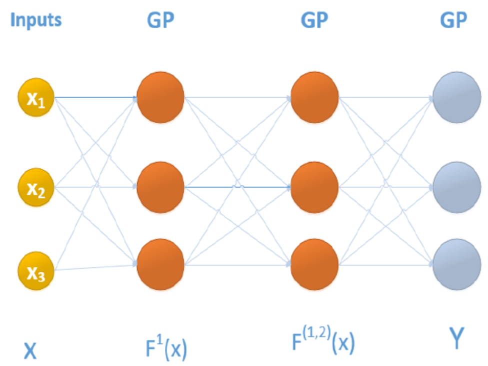


Figure 3: Example of a figure caption.

represented through a co-variance or mean function (or the kernel) which is:

$$F(x) \sim GP(m(X), (X, X')) \tag{7}$$

here $m(x)$ is the mean function.$k(x, x')$ is the covariance function for a real process $f(x)$.

for training, the Gaussian process model require a matrix $X = [x_1, x_2, x_3, \ldots.., x_n]$. the size of matrix X is usually $n * d$ and represent the data points.The model also requires a vector Y, that is $[y_1, y_2, y_3, \ldots\ldots, y_n]^T$ for training Dataset, having size $n * l$. f is usually the latent function and $f(x_i)$ is the predictive class membership probability. The Gaussian model performs mapping of latent space to the observation space which are usually non linear. The kernel function is used to change non- linear input space into a space which is represented by various dimensions in a manner where the output of the problem can be shown in a linear manner. The major problem of this model lies when a probability for a given test point$X *$ is to be predicted. Where$X *$ is a member of one of the mentioned classes. The positive class membership probability$p(y = C + |x)$ is represented by using the following equation: p (y = C+ | x) = sig (f (x))

$$p(y = C + |X) = sig(f(x)) \tag{8}$$

where f is a latent function f:$R- > R$ which is mapped into the interval[0, 1] through a sigmoid function such that: $sig : R- > [0, 1]$.

hence, sig is a sigmoid transformation function. x is a data point belonging to the finite Dataset.

The classification through classifier is made using supervised learning, hence, every data point also contains a class label: $Y_i \in C+, C-$ where $C+$ belongs to positive class and $C-$ belongs to the negative class.

3.4. Accuracy of Gaussian and Random Forest Model

For comparing the results obtained through Gaussian and Random Forest model, the accuracy of predicted encounters is calculated at different locations through both approaches. The assessments of accuracies of predictions made by both models were done using the Mean Absolute Error (MAE) of encounter prediction. The accuracy was obtained by comparing observed encounters with the predicted ones. In equation 7, MAE obtained the accuracies of predictive models by calculating how close each prediction ratio was arrayed around the original data.

$$MAE = \frac{(\sum_{n=i}^{n} |y_i - x_i|)}{n} \tag{9}$$

where x_i is the encounter prediction made for each location, and y_i is the target encounter value and n is the total number of locations for which encounters are predicted?

3.5. Scenario for the Random Forest Model Encounter Prediction For energy

Random forest model works on the principle of decision trees. Decision trees possess the decision making ability based on the results of a specific condition. Figure 1, 2 shows the sample decision trees that we have used in our work for storing history of mobility traces of three users and for making predictions about their future possible encounter. For example, Alice visits mostly cafeteria first at a specific day and time and then the library at a particular time. Similarly, Bob and Jane also have their mobility traces that are stored in the decision tree. Each location that the user visits and the specific time at which the location is visited is also stored in the decision tree. The separate storage of each spatial-temporal pair of different users, makes the retrieval of history data and predictions of encounter very easy. The average of future location prediction of each user results in the encounter prediction by the random forest model.

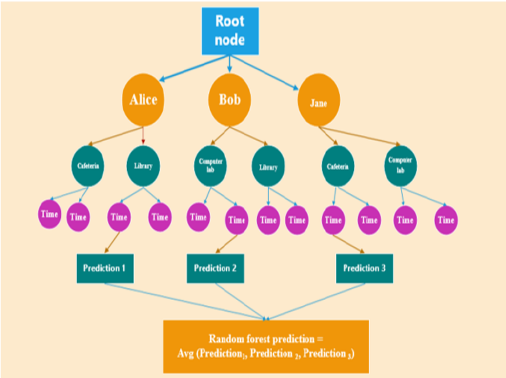


Figure 4: Encounter prediction by the random forest model.

3.6. Energy consumption model for sustainable energy approach

We have a network model to represent our problem to calculate the more sustainable energy approach. Let us assume $G = (N, L)$ be directed multiple networks with N as a number of nodes and L is all set of set of edges in the network. In a network, edges represent the flow of object from i to j between n nodes is$(i, j), i, j \epsilon N$. Edges has $(i, j) \epsilon E$ a capacity and a function of energy consumption $Ci, j, S \epsilon N$. Show the sender node and transshipment vertex $x \epsilon N$. Bandwidth also based on network link capacity, and they defined as $u_i j$. We consider demand and source $B_i < 0$, in between a relay node with demand $B_i = 0$ and a demand destination $B_i > 0$.

For prototyping, assumed a general and simplified energy consumption model for wireless or wired energy dissipation where the transmitter dissipates power to generate radio or line electronics, the power amplifier consumes energy to transmit traffic, and the receiver dissipates energy to receive and process radio or line electronics.

Taking radio transmission as an example, by setting the power amplifier accordingly, the power control can be used to correct the loss of signal propagation. For example, if the transmission distance is less than the threshold d_0, the free space prorogation model is used with the attenuation parameter ϵ_{fx}, otherwise, the multi path (mp) propagation model is used with the attenuation parameter $\epsilon_m p$:

$$E_{TX}(k,d) = E_{TX} - e_{lec}(k) + E_{TX} - e_{lec}(k,d) \tag{10}$$

For the case of radio transmission,

$$E_{TX}(k,d) = \begin{cases} k.E_{elec} + K.\epsilon_{fx}.d^2, d < d_0 \\ k.E_{elec} + K.epsilon_{mp}.d^4, d > d_0 \end{cases} \tag{11}$$

Where E_{elec} is the energy consumed by the transmitter, it depends on variables such as the processing of digital coding, modulation, and filtering signal. As for the power used by the amplifier, the distance to the receiver and the acceptable bit error rate depends on it. It is possible to measure the energy consumption of the data obtained by:

$$E_{Rx}(k) = E_{Rx-elec}(k) = kE_{elec} \tag{12}$$

Based on the above general energy consumption model for communications, two significant and changeable factors will vary the overall energy consumption compared to the energy consumed by the electronic components and also the signal processing mechanism in the transmitter, receiver, and also the relay amplifiers, as can be seen in the amount of data k and the transmission distance. Can the transmission distance d that is being traversed through the networks be reduced? We explain the utilization of the minimum cost flow problem for the different number of data volume and assess less energy consumption by transferring the data between two location and nodes.

3.7. Dataset

For evaluating the proposed work, we used Dataset of USC university of Southern California (24) traces. These traces are real time mobility measurements taken from the USC.

	Time	IP_Address	Some_value	MAC_Address	Month
1226	2004 Jan 1 00:03:22	1	5	xx:xx:xx:62:1d:60	Jan
1227	2004 Jan 1 00:03:57	1	4	xx:xx:xx:62:1d:60	Jan
1228	2004 Jan 1 00:07:48	1	5	xx:xx:xx:62:1d:60	Jan
1229	2004 Jan 1 00:10:03	1	4	xx:xx:xx:62:1d:609	Jan
1230	2004 Jan 1 00:20:39	1	5	xx:xx:xx:62:1d:60	Jan
...	...	...	...	...	...
726336	2004 Jan 31 23:51:25	2	5	xx:xx:xx:65:c:21	Jan
726337	2004 Jan 31 23:53:35	2	5	xx:xx:xx:26:c1:48	Jan
726338	2004 Jan 31 23:55:05	2	3	xx:xx:xx:65:c:21	Jan
726339	2004 Jan 31 23:56:31	2	5	xx:xx:xx:19:4e:24	Jan
726340	2004 Jan 31 23:58:50	3	15	xx:xx:xx:18:79:e3	Jan

Figure 5: Instance of dataset

4. RFEPM Numerical studies

In this section, we have evaluated the proposed work for encounter prediction among users. The evaluation is done step by step, at first, the results for encounter were calculated using both Gaussian and Random forest approach. After this, we computed encounter prediction results with both approaches and a comparison among the results obtained through both approaches was performed.At the end calculate the energy based on RFEPM results.

4.1. Results of encounter prediction through Gaussian process

These results were calculated for five users for increasing the efficiency. In results we considered two variables i.e. day of the week and hour of the day. If an encounter was observed at that hour we considered it as 1 and as 0 in other case. For calculating the encounters, we considered the weeks at which the encounter occurred at the specified day and the hour.

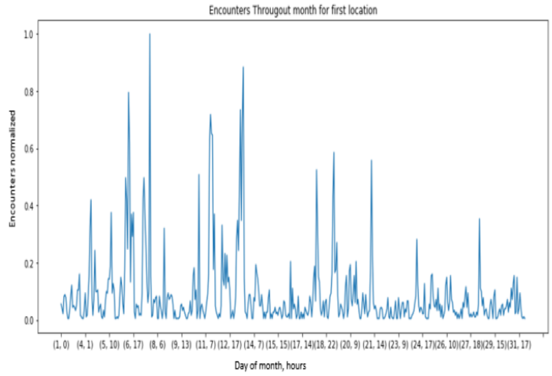


Figure 6: Encounters observed in Cafeteria.

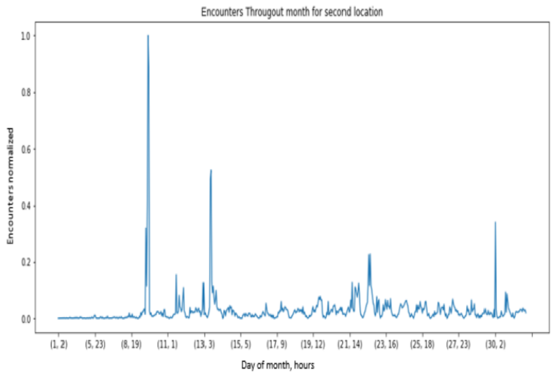


Figure 7: Encounters observed in Library .

Fig 6, Fig 7 represent the encounters that occurred in the month of January at the cafeteria and library . These encounters were actual observed encounters. The x-axis shows the (day, encounters occurred) ordered pair of the month at which the encounter occurred. Y- axis shows the normalized values of the encounters.

The encounters in figure 6, mentioned above, occurred in the month of January at cafeteria. the highest normalized values of the encounters were on the day 6 and day 24 of the month January. As on both these days 17 encounters occurred. The lowest number of encounters were on day 1, as no encounter occurred at that day in cafeteria. variable lengths of the peaks show the different values of encounters that occurred on different days. Similarly, in computer lab i.e. Figure 7, the highest number of encounters occurred at day 8 which are 19. The lowest number of encounters were on day 11, as only one encounter occurred at that day in computer lab. Figure 8,9. shows the encounter prediction vs actual encounter observed through Guassian process

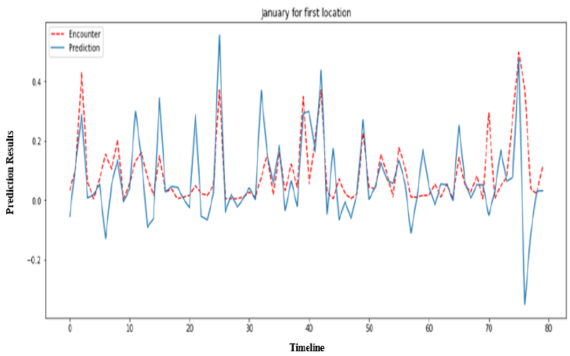


Figure 8: Show the encounter prediction results through Gaussian model for Cafeteria.

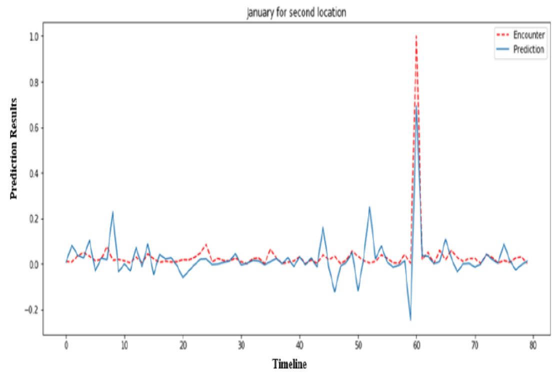


Figure 9: show the encounter prediction results through Gaussian model for Library.

4.2. Results of encounter prediction through RFEPM

Figure 10,11 shows the encounter prediction vs actual encounter observed through RFEPM at location 1 and location 2. The x-axis shows the total number of weeks in the whole Dataset when encounters occurred. Y-axis shows the normalized values of encounters.

4.3. Accuracy of Gaussian and Random Forest Model

For comparing the results obtained through Gaussian and Random Forest model, we calculated accuracy of predicted encounters at different locations through both approaches. These results compared accuracy of predicted encounter at a specific location in different weeks of the month.

The graphical representation of accuracy comparison for results obtained through both approaches (Gaussian and Random Forest) is shown in figure. An evident higher accuracy of Random Forest model can be seen in the figure 12. The highest accuracy achieved by Random forest model was for location 5 which was more than 80% .

4.4. Energy consumption base on RFEPM

After getting the RFEPM model we know where and which time people meet so basis on the RFEPM we calculate the 200 users select the Dataset randomly and as we calculate the D2D and internet and show that D2D is better than internet. (24) Although Figure 13 shows energy consumption increases with the number of users transmitting data in the network but still it is low as compared to that consumed for data transfer using internet. We contrasted and evaluated

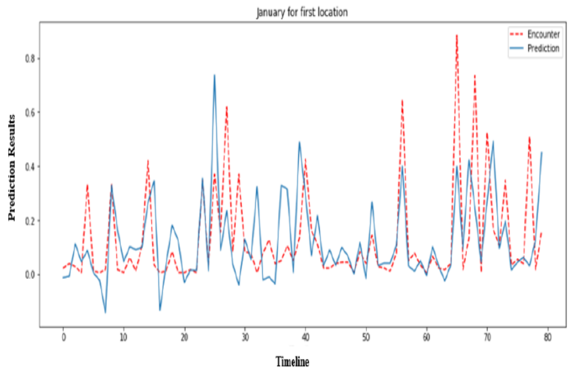


Figure 10: Show the encounter prediction results through random forest model for Cafeteria.

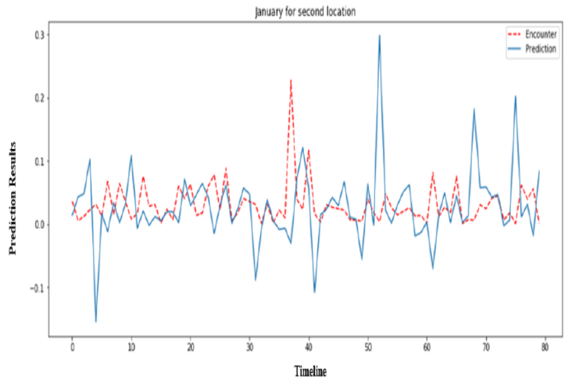


Figure 11: Show the encounter prediction results through random forest model for Library.

the Gaussian process and random Forest use data traces of real-world mobility and demonstrating how these effects of encounter predictions save resources. The purpose of our proposed model is to reduce the resource consumption i.e. energy during data transmission in daily movements .Our RFEPM method can save the energy using Random Forest Model in smart connected buildings.

5. CONCLUSION

We have used human mobility patterns for an efficient, ensuring, and less resource (energy, bandwidth) consuming data transmission process. Our proposed model knows energy use about occupant behavioral trends from the building. We also work on encounter prediction i.e. predicting when two or more users will encounter at a specific place and time. The exact encounter prediction can help to avoid resource consumption that is used continuously searching for any available encounter. For encounter prediction, we used random forest model that stored the history of human mobility traces in tree form. For encounter prediction, the model at first predicts the next future location. Then based on the similar mobility preferences an encounter is predicted at a specific day of the month and hour of the day. We have also provided a mathematical computational model to encounter the prediction of different users. For experimentation, we have trained the USC dataset with Benchmark Gaussian and Random forest model. The results were computed for observed and predicted encounters through both approaches. The accuracy of predicted encounters and their comparison were also shown in the paper. The results showed that the accuracy of encounter prediction

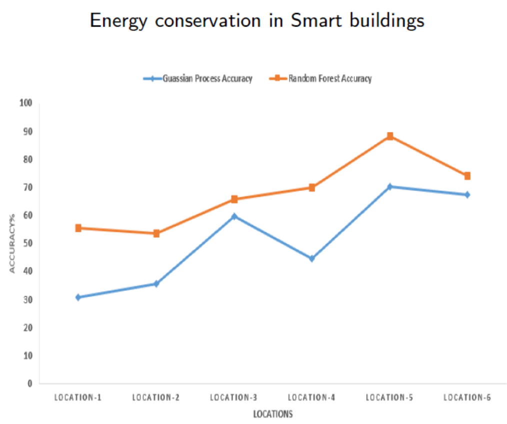


Figure 12: Graphical representation of comparison among accuracy of Gaussian and Random forest model

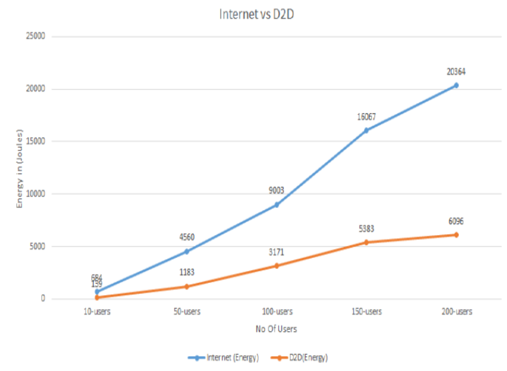


Figure 13: Energy consumption based on Internet and D2D.

through the random forest model was quite higher than Benchmark Gaussian model. Even at some locations, the encounters were predicted even with an accuracy higher than 80%

Our proposed method can save the energy using Random Forest Model in smart connected buildings. In the future, we further plan to investigate energy-saving models.

References

[1] K. Khalid, I. Woungang, S. K. Dhurandher, L. Barolli, G. H. Carvalho, and M. Takizawa, "An energy-efficient routing protocol for infrastructure-less opportunistic networks," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2016 10th International Conference on*. IEEE, 2016, pp. 237–244.

[2] V. Stavroulaki, K. Tsagkaris, M. Logothetis, A. Georgakopoulos, P. Demestichas, J. Gebert, and M. Filo, "Opportunistic networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 3, pp. 52–59, 2011.

[3] S. K. Dhurandher, D. K. Sharma, and I. Woungang, "Energy-based performance evaluation of various routing protocols in infrastructure-less opportunistic networks." *J. Internet Serv. Inf. Secur.*, vol. 3, no. 1/2, pp. 37–48, 2013.

[4] A. Vahdat, D. Becker *et al.*, "Epidemic routing for partially connected ad hoc networks," 2000.

[5] A. Lindgren, A. Doria, and O. Schelén, "Probabilistic routing in intermittently connected networks," *ACM SIGMOBILE mobile computing and communications review*, vol. 7, no. 3, pp. 19–20, 2003.

[6] T.-K. Huang, C.-K. Lee, and L.-J. Chen, "Prophet+: An adaptive prophet-based routing protocol for opportunistic network," in *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*. IEEE, 2010, pp. 112–119.

[7] C. Boldrini, M. Conti, J. Jacopini, and A. Passarella, "Hibop: a history based routing protocol for opportunistic networks," in *World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a*. IEEE, 2007, pp. 1–12.
[8] T. Kathiravelu, N. Ranasinghe, and A. Pears, "A robust proactive routing protocol for intermittently connected opportunistic networks," in *Wireless And Optical Communications Networks (WOCN), 2010 Seventh International Conference On*. IEEE, 2010, pp. 1–6.
[9] J. Makhlouta, H. Harkous, F. Hutayt, and H. Artail, "Adaptive fuzzy spray and wait: efficient routing for opportunistic networks," in *Mobile and Wireless Networking (iCOST), 2011 International Conference on Selected Topics in*. IEEE, 2011, pp. 64–69.
[10] J. Han, E. Owusu, L. T. Nguyen, A. Perrig, and J. Zhang, "Accomplice: Location inference using accelerometers on smartphones," in *Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on*. IEEE, 2012, pp. 1–9.
[11] H. Holbach and M. A. Bourassa, "The effects of gap wind induced vorticity, the itcz, and monsoon trough on tropical cyclogenesis," in *2012 International Ocean Vector Winds Science Team Meeting*, 2012.
[12] F. Calabrese, G. Di Lorenzo, and C. Ratti, "Human mobility prediction based on individual and collective geographical preferences," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, 2010, pp. 312–317.
[13] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 637–646.
[14] V. Bellotti, B. Begole, E. H. Chi, N. Ducheneaut, J. Fang, E. Isaacs, T. King, M. W. Newman, K. Partridge, B. Price *et al.*, "Activity-based serendipitous recommendations with the magitti mobile leisure guide," in *Proceedings of the sigchi conference on human factors in computing systems*. ACM, 2008, pp. 1157–1166.
[15] L. Vu, Q. Do, and K. Nahrstedt, "Jyotish: A novel framework for constructing predictive model of people movement from joint wifi/bluetooth trace," in *Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on*. IEEE, 2011, pp. 54–62.
[16] G. Yang, "Discovering significant places from mobile phones–a mass market solution," *Mobile Entity Localization and Tracking in GPS-less Environnments*, pp. 34–49, 2009.
[17] D. Ashbrook and T. Starner, "Learning significant locations and predicting user movement with gps," in *Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on*. IEEE, 2002, pp. 101–108.
[18] R. Chellapa, A. Jennings, and N. Shenoy, "A comparative study of mobility prediction in fixed wireless networks and mobile ad hoc networks," in *Communications, 2003. ICC'03. IEEE International Conference on*, vol. 2. IEEE, 2003, pp. 891–895.
[19] G. Yang, "Discovering significant places from mobile phones–a mass market solution," *Mobile Entity Localization and Tracking in GPS-less Environnments*, pp. 34–49, 2009.
[20] L. Vu, Q. Do, and K. Nahrstedt, "Exploiting jointwifi/bluetooth trace to predict people movement," 2010.
[21] R.L. Plackett, "Big data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook,"."
[22] R. Ciobanu and Dobre, "High performance mobile opportunistic systems," 2012, October.
[23] e. a. S. Eisele, Scott, "A resilient and trustworthy transactive platform for smart and connected communities." 2018 ieee international conference on internet of things (ithings) and ieee green computing and communications (greencom) and ieee cyber, physical and social computing (cpscom) and ieee smart data (smartdata). ieee, 2018," 2018.
[24] Marincic and I. Foste, ""energy-efficient data trans- fer: Bits vs. atoms," in software, telecommunications and computer networks," in *(SoftCOM),*, 2016 24th.