UNIVERSITY OF LEEDS

Green Vehicular Content Distribution Network

Samaneh Igder

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

School of Electronic and Electrical Engineering

University of Leeds

February 2017

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

To:

My dearest parents, Fatemeh & Kamran

Acknowledgement

I could not have completed this journey if this was not for the help and encouragement of my dear supervisor, Professor Jaafar Elmirghani. I am grateful for his teachings, guidance and support. I would like to sincerely acknowledge his help and support.

I would also like to use this opportunity to thank my friends and colleagues in the Communication Systems and Networks group, who have been instrumental in the success of my research project. I send my special thanks to Dr. Bilal Qazi and Dr. Samya Bhattacharya, Postdoctoral Research Fellows in the group, for their useful advice and fruitful discussions.

I am blessed with a warm and loving family. My dear parents have done more for me during the course of my life than words can describe. I cannot wait to see their proud faces along with my sisters on my graduation day and hope I will have the opportunity to express my gratitude towards them for many years to come. I am indebted to my mother for her patience, unconditional support, and encouragement. Many thanks to my sisters who offered invaluable support over the years and special thanks to Dr Raúl Hernández Aquino who fully supported me and was always there during difficult times, especially considering he was going through the same journey.

This is also an appropriate venue to thank Lorraine Stokes and her lovely family who have been a source of energy and created a warm environment for me.

Thanks for supporting me during my studies and urging me on. In the early stages of my PhD, I lost my dearest uncle. He was so incredibly dear to me and his loss has left a hole in my heart forever. I believe every success in my life was partly thanks to his constant prayers and this PhD is no exception.

Finally, to all of you! Thanks for always being there for me.

Abstract

With environmental awareness becoming a global concern, content distribution has become popular in the context of modern city scenario with obvious concerns for ICT power consumption. The business world demands huge amounts of information exchange for advertisement and connectivity, which is an integral part of a smart city.

In this thesis, a number of energy saving and performance improvement techniques are proposed for the content delivery scenario. These are: content cache location optimisation techniques for energy saving and transceiver load adaptive techniques that save energy while maintaining acceptable piece delay. With the recent advancement in Fog computing, nano-servers are introduced in the later part of the thesis for content delivery and process of user demands. Two techniques random sleep cycles and rate adaptation are proposed to save transmission energy. The quality of service in terms of piece delay and dropping probability are optimised by deploying renewable and non-renewable energy powered nano-servers (NS). Finally, mixed integer linear programming models (MILP) were developed alongside other optimisations methods like bisection, greedy and genetic algorithms which judiciously distribute renewable energy to the fog servers in order to minimise the piece delay and dropping probability in heavily loaded regions of the city area.

Contents

List of Figures

List of Tables

List of Abbreviations

AP	Access Point
BS	Base Station
C2CC	cars are Car-to-Car Communications
CCDSV	Cooperative Content Distribution System for Vehicles
CDN	Content Distribution Network
CH	Cluster Head
CP	Cache Points
CS	Candidate Site
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CuTV	Catch-up TV
DCH	Double Cluster-Head
DSRC	Dedicated Short Range Communication
FIFO	First-In-First-Out
GPS	Global Positioning System
GSM	Global System for Mobile communication
HVC	hybrid vehicular communication
ICT	Information and Communication Technology

IoT	Internet of Things
IoTCD	IoT Content Delivery
IP	Internet Protocol
IPTV	Internet Protocol TV
ITS	Intelligent Transportation System
IVC	Inter-Vehicle Communication
MAC	Medium Access Control
MANET	Mobile Ad Hoc Network
MH	Multi Hop
MILP	Mixed Integer Linear Programming
MIVC	Multi-hop Inter-Vehicle Communication
MPR	Market Penetration Ratio
NS	Nano Server
OBU	On-board unit
OFDM	Orthogonal Frequency Division Multiplexing
ONU	Optical Network Unit
PCD	Push-Based Popular Content Distribution
pdf	Probability Density Function

QoS	Quality of Service
RA	Rate Adaptation
RDP	Request Dropping Probability
RSU	Roadside Unit
RVC	roadside to vehicle communication
SC	Sleep Cycles
SLA	service level agreement
SLNC	Symbol Level Network Coding
SV2R	Sparse Vehicle-to-Roadside
TCP	Transmission Control Protocol
TP	Traffic Point
UMTS	Universal Mobile Telecommunications System
V2R	Vehicle-to-Roadside
V2V	Vehicle-to-Vehicle
VANET	Vehicular Ad Hoc Network
VCDA	Vehicular Content Download Algorithm
VoD	Video-on-Demand
VSC	Vehicle Safety Communication

WAVE	Wireless Access in Vehicular Environments
WLAN	Wireless Local Area Network
WSMP	WAVE short message protocol
WWAN	Wireless Wide Area Network

List of Symbols

TP	Set of traffic points
CP	Set of caching point candidates
T	Set of time points within one hour i.e. 600
$Pmax_{ct}$	Maximum power consumption of of $CP\ c$ at time t i.e. $20\ W$
$Pidle_{ct}$	Operational power consumption of $CP\ c$ at time t, $Ceiling(P_{ct}\ /\ (1.3548) = 14\ W$
$Ptxmax_{ct}$	Maximum transmitter power consumption of CP c at time t , P_{ct} - $Pidle_{ct}$=6 W
$Ptxmin_{ct}$	Minimum transmitter power consumption of $CP\ c$ at time t i.e. $1\ W$
Pw	Hourly available wind power
λ_{jt}	Traffic demand at $TP\ j$ time t
$Ndmax_{ct}$	Maximum number of simultaneous downloads of $CP\ c$ at time t (i.e. 10)
$drmax_{ct}$	Maximum Data rate at $CP\ c$ at time t (i.e. 30 Mbps)

$drmin_{ct}$	Minimum Data rate at $CP\ c$ at time *t* (3 Mbps)
A	Large number for MILP constraint, set to 600 here
P_{ct}	Power consumption of a CP i.e. $14{\sim}20\ W$
Ptx_{ct}	Power transmitting consumption of CP c at time t
Pre_{ct}	The part of the available renewable used by CP c at time t
$Pnre_{ct}$	Non-renewable part of required power consumption of $CP\ c$ at time t
λ_{cjt}	Traffic between $CP\ c$ and $TP\ j$ at time t
dr_{ct}	Adaptive Data rate at $CP\ c$ at time *t* (3~30 Mbps)
α_c	Equals 1 if $CP\ c$ is on, equals 0 otherwise
β_{ct}	Equals 0 if Renewable energy is adequate for power consumption of CP, equals 1 otherwise
δ_{cjt}	Equals 1 if $CP\ c$ is transmitting content to Vehicle *j*, equals 0 otherwise
$Ndmax$	Maximum number of simultaneous downloads from a NS (9)

Nd_{ct}	Adaptive number of simultaneous downloads of $CP\ c$ at time t (i.e. $0{\sim}10$)
NS	Set of installed Fog computing Nano servers
$N[j]$	Set of neighbouring TPs for $NS\ n$
$N[n]$	Set of neighbouring NSs for $TPs\ j$
$Bmax$	Maximum capacity of a NS (27 Mbps)
$drmax_{nt}$	Maximum data rate at $NS\ s$ at time t (i.e. 27 Mbps)
$Dmax_n$	Maximum acceptable average piece delay at $NS\ s$ at time t (i.e.13 s)
$Pmax_{nt}$	Maximum power consumption of at $NS\ s$ at time t (30 W)
$Pidle_{nt}$	Operational power consumption of $NS\ s$at time t 23 W
$Ptxmax_{nt}$	Maximum transmission power consumption of a NS at time t given by $Pmax_{nt} - Pidle_{nt} = 7\ W$
Pw	Hourly available wind power
$Pmin_{nt}$	Minimum required power by $NS\ n$ at time t
$Esmax_n^s$	Maximum energy saving of $NS\ n$ with random sleep

	cycles
$Psmax_n^s$	Maximum power saving of $NS\ n$ with random sleep cycles
Ewo	Wake-up overhead energy in Joule
λ_{jt}	Traffic demand at $TP\ j$ at time t
$\overline{Smax}_n$	Maximum sleep duration at $NS\ n$
σ	A constant, set to 10
Ptx_{nt}	Adaptive transmission power of $NS\ n$ at time t
$Ptxmin_{nt}$	Minimum transmission power of $NS\ n$ at time t according maximum acceptable delay
Es_n	Energy saving of $NS\ n$
Es_n^s	Energy saving of $NS\ n$ with random sleep cycles
Es_n^{ra}	Energy saving of $NS\ n$ with rate adaptation
Ps_n^s	Power saving of $NS\ n\ t$ with random sleep cycles
Ps_n^{ra}	Power saving of $NS\ n$ with rate adaptation
α_n	Equals 1 if $NS\ n$ is on, equals 0 otherwise
λ_{njt}	Traffic between $NS\ n$ and TP j at time t

λ_n	Traffic demand at $NS\ n$ time t
δ_{njt}	Equals 1 if $NS\ n$ is transmitting IoT content to $TP\ j$, equals 0 otherwise
Pre_{nt}	The part of the available renewable used by $NS\ n$ at time t
$drmax_{nt}$	Maximum data rate at $NS\ s$ at time t (i.e. 27 Mbps)
$Dmin_n$	Minimum average piece delay at $NS\ s$ at time t(when data rate is $drmax_{nt}$)
dr_{nt}	Adaptive data rate at $NS\ n$ at time t
$\bar{S}_n$	Sleep duration at $NS\ n$
$\overline{Smax}_n$	Maximum Sleep duration at $NS\ n$
NOs_n	Number of sleep cycles at $NS\ n$
Dra_n	Average piece delay at the rate adaptive $NS\ n$
Dsc_n	Average piece delay at $NS\ n$ with sleep cycles
RDP	Maximum acceptable request dropping probability (i.e. 0.05)
$Ndmax$	Maximum number of simultaneous downloads from a

	NS (9)
P_{nt}	Adaptive power consumption by $NS\ n$ at time t
R_{jt}	Request data rate at $TP\ j$ at time t
Ptx_{nt}	Adaptive transmission power of $NS\ n$ at time t
$Ptxmin_{nt}$	Minimum transmission power of $NS\ n$ at time t according maximum acceptable delay
R_{njt}	Request data rate between $NS\ n$ and TP j at time t
R_n	Total Request data rate at $NS\ n$ at time t
δ_{njt}	Equals 1 if $NS\ n$ is transmitting IoT content to $TP\ j$, equals 0 otherwise
Pre_{nt}	The part of the available renewable used by $NS\ n$ at time t

Chapter 1: Introduction

In the modern age, smart cities have become popular and essential to the city dwellers due to the fast paced life. The main part of being “smart” is utilising information and communications technology (ICT) and the Internet to address civil challenges. The majority of the world’s population is living in urban areas, by 2050 more than 66 percent of the world’s population will be living in cities [1,2]. The resident populations of cities are growing by nearly 60 million every year [3]. Managing urban areas has become one of the most important development challenges of the 21st century. Transportation systems are considered as a critical element in our city lives. Vehicular communication as a key future technology in intelligent transportation systems is essential in smart cities. The infotainment data is dynamic in nature. Thus, data is largely affected by mobility, number of users and real time applications. To deal with such transiency, the radio range needs to have high bandwidth (and hence shorter distance). With the latest advances in the Internet of Things (IoT), a new era has emerged in the smart city domain [1], opening new opportunities for the development of efficient and low-cost applications that aim to improve the quality of life in cities. To solve such issue, intermediate devices (between cloud and end users) are needed [2]. The abovementioned reasons motivate this study.

The research objectives of this thesis were:

- To study the topology and characteristics of a realistic city vehicular communication environment in order to develop an accurate vehicular simulator that captures the city vehicular dynamics.
- To introduce load adaptive cache points and evaluate the performance of energy efficient city vehicular communications scenario.
- To introduce and study the performance of *energy adaptive nano-servers* in the Internet of Things (IoT) information delivery in smart city vehicular environment and optimise the overall energy consumption.
- To investigate the feasibility of energy saving by introducing sleep mechanism while maintaining an acceptable latency.
- To investigate the feasibility of further energy saving by considering request dropping which results in non-zero *Request dropping Probability (RDP).*
- To introduce renewable energy for further reduction in non-renewable energy consumption and its judicious distribution in the city in order to optimise performance.

1.1 Key Contributions

1. Developed a load adaptation technique for CPs which not only enhances content download rate but also reduces transmission energy

consumption through random sleep cycles. Each CP with a wireless link to vehicles is modelled as a single server queue (M/det-Neg/1/∞). The performance of the system was evaluated in terms of energy savings, average queue size, average piece delay, and average content delay. The performance results of the system are verified through simulations with respect to varying vehicular load according to real vehicular traffic profiles.

2. Studied vehicular traffic measurements of a typical city environment such as speed, acceleration, deceleration, and inter-vehicle spacing and developed an accurate city vehicular simulator based on a $3 \times 3\ \mathrm{km}^2$ Manhattan grid which captures the mobility and dynamics of vehicular motion.
3. Developed vehicular traffic simulator to understand the mobility and traffic characteristics of vehicles in a smart city network, and measured vehicular traffic flow profiles.
4. Developed a Mixed Integer Linear Programming model (MILP) for the introduced renewable energy grid (in the form of wind energy) and adaptive CPs location optimisation to reduce non-renewable energy consumption (carbon footprint) and further, proposed algorithms for provisioning high number of simultaneous downloads, which reduce overall waiting time and number of dropped request of city vehicular users. The performance of this model was evaluated in terms of the number of active caching points, total network energy consumption,

energy saving, Number of simultaneous downloads and average piece delay. The performance results of the system are verified with developed heuristic model.

5. Studied the advantage of the Internet of Things (IoT) in smart cities cloud computing and intermediate devices fog (between cloud end users and fog nano-servers as a development of efficient and low-cost applications solution for transiency and rapid growth of users. Proposed energy efficient adaptive nano-servers for IoT contents delivery and developed five different MILP models:
    a. Model 1 minimises the power consumption of IoTCD by optimising the number and location of the Fog Servers (FSs) while accounting for vehicle mobility in the Fog Servers for IoT content delivery (IoTCD).
    b. Model 2 reduces the non-renewable transmission power consumption of the IoTCD by introducing random sleep cycles while maintaining the minimum acceptable QoS.
    c. Model 3 reduces the non-renewable transmission power consumption of the IoTCD by introducing energy aware rate adaptation while maintaining the minimum acceptable QoS.
    d. Model 4 improves the overall network QoS and reduces carbon footprint by optimally distributing the available renewable energy according to the IoT Contents demand at each FS, hence reducing waiting delay.

e. Developed MILP models for energy adaptive nano-servers for IoT Information Piece Delivery (IoTIPD).

1.2 Publications

1. “Load Adaptive Caching Points for a Content Distribution Network”, Samaneh Igder, Hamdi Idjmayyel, Bilal R. Qazi, Samya Bhattacharya, Jaafar M. H. Elmirghani, in Proc. NGMAST2015, UK.
2. “Standalone Green Cache Points for Vehicular Content Distribution Networks”, Samaneh Igder, Samya Bhattacharya, Bilal R. Qazi, Jaafar M. H. Elmirghani, accepted in NGMAST2016, UK.
3. “Energy Efficient Fog Servers for Internet of Things Information Piece Delivery (IoTIPD) in a Smart City Vehicular Environment”, Samaneh Igder, Samya Bhattacharya, Jaafar M. H. Elmirghani, accepted in NGMAST2016, UK.
4. “Energy Efficient Fog Computing in Smart City Vehicular Environment”, Samaneh Igder, Samya Bhattacharya, Bilal R. Qazi, Jaafar M. H. Elmirghani, submitted to IEEE Transactions on Vehicular Technology.

1.3 Thesis Outline

Following the introduction, the thesis is organised as follows:

Chapter 2 summarises vehicular network applications, architectures, technologies and environments. The chapter also presents an overview of Wireless Access in Vehicular Environments (WAVE) standard and Cloud and Fog Computing for Internet of Things for vehicular networks.

Chapter 3 describes single cache point, which adapts its data rate according to demand and goes to sleep when the demand is low.

Chapter 4 introduces mixed integer linear programming optimisation model to minimise power consumption while serving traffic.

Chapter 5 As the PhD progressed, fog computing became as important as content caching. Therefore, nano-servers are introduced in this chapter for processing as well as content delivery along with three optimisation technique to verify MILP.

Chapter 6 introduces transmission energy saving through random sleep cycles and rate adaptation. Since these techniques affect delay and if more power is available, the potential reduction in delay is investigated.

Chapter 7 explores another way of power saving, where some requests are deliberately dropped while remaining within the acceptable service level

agreement (SLA). The performance optimisation using the dropping as a measure is investigated. The performance of the system is evaluated in terms of QoS and energy savings and the results are verified by genetic algorithms.

Chapter 8 summarises the major contributions of this thesis and highlights the recommendations for future investigations.

Chapter 2: City Vehicular Environment

2.1 Introduction

One of the high priority research areas in present times is the development of road transport communication, as roads are one of the major modes of transportation in the modern era. Considering the importance of both safety and economic concerns, this area of research has witnesses growing importance recently [3]. The main concern in this field is to ensure safe and comfortable travelling as well as the shortest possible travelling time [4]. As the demand for and usage of vehicles in today's world increases, traffic congestion, delay and accidents in the transportation system occur more frequently which in turn lead to significant expenditure and energy resources and the loss of life [5]. In order to improve the safety and security of the transportation system, existing technologies face considerable challenges [6-8].

Today, vehicles are increasingly equipped with on-board computing and communication devices. A brief look at the past reveals that the approach of using wireless communication in vehicles started around the 1980s [9]. While car phones, on-board Internet and Bluetooth are existing examples of wireless communications technologies in modern vehicles, today, corporations and

scientists are aiming towards the implementation of preventive systems, which are able to foresee dangerous situation on the roads. The prospect of detecting an un-safe situation before it actually occurs is the focus of many projects in the US, Europe and Japan. Furthermore, the decreasing cost and rapid growth of wireless technologies has motivated research in this area. Since then, various Intelligent Transportation Systems (ITS) [10], including car navigation systems [11], traffic signal control systems, automated number plate recognition [12], speed cameras, parking guidance and information systems, weather information systems, and so on, have come under focus. More recently, however, more advanced ITSs using wireless communication technology have been considered for collision avoidance and advanced detection of danger [13]. ITS councils became involved in developing technologies and applications for vehicular communication networks [14]. The aforementioned technologies contribute to the safety and comfort in the world of driving. The topic of Inter-Vehicle Communication has attracted more and more attention in the last few years [15], since it can make transportation secure and effective for drivers, passengers and pedestrians [16].

Communication systems in vehicles can be classified into three major types, which include: a) infrastructure-less (Ad hoc) communication systems, b) infrastructure-based communication systems and, finally, c) Hybrid communication systems [17]. Communication in vehicular systems can in turn be categorised as:

- one–to-one, e.g. playing games, sharing pictures and video .
- one-to-many, e.g. a *roadside unit* (RSU) transmitting curve warning signal.
- Many-to-many, e.g. vehicles sharing common information (location, speed, etc.) with other vehicles within the communication range [16].

In vehicular networks, applications may be categorised as safety or non-safety related services. Vehicular ad hoc networks (VANETs) are a type of Mobile ad hoc networks (MANETs) known to have no fixed infrastructure, and rely on mobile nodes in disseminating the information throughout the network [17]. VANETs work by enabling wireless Vehicle-to-Vehicle communications (V2V), in addition to Vehicle-to-Infrastructure (V2I) communications, and vehicle-to roadside (V2R) communication systems [18]. VANETs possess an advantage over many other wireless technologies such as 3G and cellular networks in terms of packet delivery in vehicular environments. Peer vehicles within transmission range are able to exchange data packets amongst each other using a wireless radio. These packets might contain information about traffic situations, such as traffic jams, road constructions, and car accidents, which can be utilised to organise traffic and provide safety for both drivers and passengers. Moreover, VANETs are also able to provide on-board entertainment for passengers.

Data can be transmitted from a vehicle to all other vehicles in range in a single-hop fashion for safety related application. For applications which are not involved with safety, a seamless communication route can be enabled from a given source vehicle to another vehicle (the destination) in a multi-hop manner where intermediary vehicles undertake the task of relaying data. Thus, routing protocols are utterly important in V2V networks for efficient data delivery. Multi-hop routing has significant functions in areas where no infrastructure is available to support communication in single-hop [19]. It can also serve to extend the coverage of an RSU, where data from a vehicle can be transmitted to an RSU through intermediary vehicles (this is useful in cases where the target vehicle is not in communication range of the source). One of the advantages of vehicular ad hoc networks is the low latency communication which results from direct communication(wherever possible), broader coverage and the added benefit that it requires no service fee [20].

The provisioning of real time multimedia data such as audio and video to users with the required Quality of service (QoS) in vehicular networks can pose a dramatic challenge due to high-density roadways and the speed of vehicles and thus requires a robust, capable and high-speed communication protocol that can enable sharing of warning messages in real-time. Thus, developing a suitable MAC protocol that can adapt to these characteristics is considered to be one of the main challenges in the field [21]. Added to this is the fact that

VANETs are characterised by the lack of a centralised entity to control and manage transmissions of the shared medium.

Realising the importance of inter-vehicle communications (IVC), the Federal Communication Commission (FCC) in the US has allowed for the allocation of a 75 MHz band in the 5.9 GHz spectrum of Dedicated Short Range Communications (DSRC) for V2V and V2I communications [22-24]. What is more, the need for reducing the carbon footprint motivated the researchers engaged in the communication field to explore various paradigms of energy efficiency in vehicular communication networks [25-28]. This has driven the growth towards hybrid fuel, electrical vehicles [29, 30].

2.2 Vehicular communication

This section provides an overview of vehicular networks with a focus on the following subjects:

1. Architectures
2. Environments
3. Applications

2.2.1. Vehicular network architecture

Vehicular networks are short-to-medium range wireless communication networks that connect a vehicle to nearby vehicles and mobile/fixed-location

resources [16]. The nodes (vehicles) in vehicular networks are equipped with an *on-board unit* (OBU) made up of sensors, a computer, a communication transceiver and a *global positioning system* (GPS) receiver for the purposes of relaying information between vehicles and between vehicles and the RSU. Utilising a GPS receiver, each vehicle is able to track its position in time and space. Communications in vehicular networks takes place in three distinct ways. These will be discussed below.

2.2.1.1. Infrastructure-less vehicular/vehicle-to-vehicle (V2V) communication system

In this kind of system, vehicles interact in an ad hoc manner and constitute a *vehicular ad hoc network* (VANET) [7, 16] and information is exchanged only between the vehicles. The vehicle-to-vehicle (V2V) communication or inter-vehicle communication (IVC) systems are very cost efficient and the network is easily deployable. By trading information, the vehicles accumulate knowledge about their local traffic situation, aiming towards increasing the safety of drivers, passengers and pedestrians. Considering the positions of the source and destination vehicles, we can further classify IVC systems as single-hop and multi-hop IVCs (SIVCs and MIVCs) [17]. In SIVCs, the source and destination of information are within radio range of each other. Communication over expanses greater than the transmission range of a node can be accomplished by utilising MIVCs, where intermediary vehicles relay information between the source and destination nodes [31].

The IVC system has two major benefits. One being that there are virtually no service charges for a direct communication between vehicles, since no service provider is needed. The other advantage of the IVCs is that the real time communication delay is small (for SIVC) when compared to centralised systems, as there is no intermediate BS or AP. On the other hand, the market penetration ratio (MPR), network availability and scalability constitute the main complications in IVC systems [16, 17]. Information exchange is only available when there are at least two vehicles equipped with intelligent devices and within each other's communication range. Continuous network connectivity cannot be guaranteed due to network partitioning. Even when a higher market penetration is achieved, scalability becomes an issue in the fair sharing of resources [32]. Added to all this is the fact that the delay involved in end-to-end communication is a major setback of the MIVC systems [17].

2.2.1.1.1. Challenges of VANETs

As was previously noted, VANETs pose various challenges which must be taken into account when designing an IVC system and reliability, robustness and time delay are the basic constituent components of these challenges.

A. High Mobility

The high mobility of nodes is usually the first goal and challenge set forth in related literature. Since vehicles are usually moving with great speed a frequent change in the network topology and hence channel partitioning is inevitable.

B. Lack of a centralised management entity

As mentioned before, VANETs operate without an existing centralised entity to organise, coordinate and manage the exchange of information of a given shared medium. This can result in an inefficient access to the channel, thus making collision between unsynchronised packets highly probable [33].

C. Transmission power

In order to increase the throughput, the transmission power of data packets must be managed properly. Because of the trade-off involved, power cannot be unboundedly increased without restraint. What is more, the density of nodes is a significant factor in controlling the transmission power; if node density is low, power should be increased so that the packets can reach their destination. In contrast, a decrease in transmission power is necessary in highly dense environments in order to reduce the range of interference and thereby causing minimum congestion [20, 34].

D. Radio channel characteristics

In some cases VANETs might not have proper characteristics necessary for wireless communications. Multipath dissemination necessitated by the presence of obstacles is one of the issues which might end up degrading the strength and quality of the signal received. Another issue to keep in mind is fading, which is caused by the mobility of the surrounding objects and the

constant movement of the sender/receiver nodes themselves [33]. Moreover, increasing the transmission power may not help to overcome the effects of fading.

F. Hidden and exposed node

Figure 2.1 portrays a common issue of wireless Ad hoc channel access. The problem arises when two given nodes are hidden from each other (i.e. out of signal range). As is clear in Figure 2.1, the two nodes are transferring a packet to the same destination simultaneously which results in data collision.

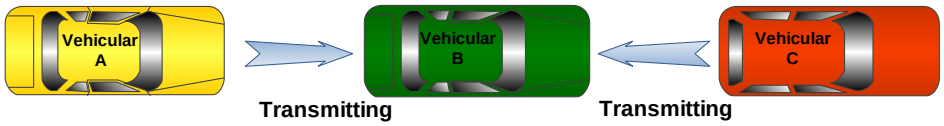


Figure 2.1: Hidden node.

If the receiver's node is to avoid collision, it should inform all of its neighbouring nodes that the channel is occupied. This can be done by transmitting a control message using a handshake protocol, as is depicted in Figure 2.2.

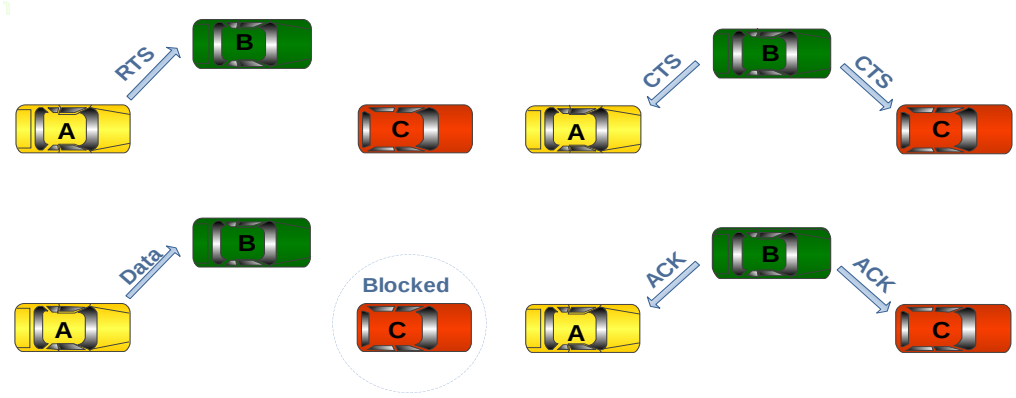


Figure 2.2: Hand Shake protocol.

The process is as follows: First, the node wishing to transmit a data packet will send an RTS (Request to Send) message to the receiver's node [shown in figure 2.2]. Once the RTS is received by the receiver node, it responds by sending back a CTS (Clear to Send) message to all of its neighbours. In the last stage it is crucial that all the neighbouring nodes, both the sender and receiver, get informed that the medium is busy so that end collisions can be avoided. In contrast to the hidden node, an exposed node is a node in range of the transmitter, but out of the receiver's range. Figure 2.3 illustrates a case that helps clarify the exposed node problem.

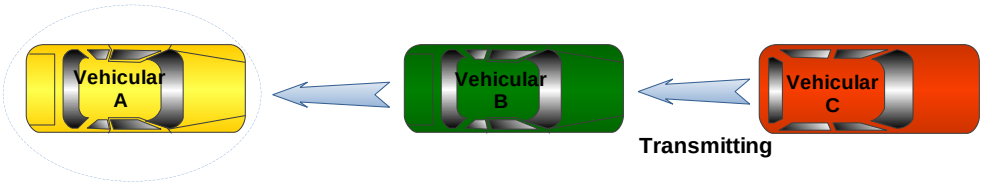


Figure 2.3 : Exposed Node Problem.

As is evident in the Figure 2.3, while node B is sending data packets to node C, node A overhears the transmission and thereby gets blocked from any other communication; this in turn decreases the efficiency of the bandwidth utilisation.

A potential solution for the aforementioned issue is the use of distinct control and data channels or the use of directional antennas.

G. Economic issue

Equipping Cars with an IVC system is an issue because consumers should be satisfied enough to buy this technology. For car to car communications, it is estimated that at least a penetration of 10% equipment rate is needed to make the network usable [35]. If half of the newly produced cars are Car-to-Car Communications (C2CC) enabled, we need about 3 years to reach 10%. On the other hand, after around 2 years company cars are resold so the respective owner before benefiting from this technology may resell their cars.

2.2.1.2 Infrastructure-based vehicular/vehicle to roadside (V2R) communication system

Ad hoc communication systems in vehicular networks cannot guarantee reliable communication. This is mainly due to the unreliable nature of V2V communication, especially in areas with low vehicular density. In addition to this, a pure V2V network does not allow access to any external online resources such as the Internet. on the contrary, infrastructure based inter-vehicular or vehicle to roadside (V2R) communication systems can be utilised for all types of communication from the vehicle to a fixed infrastructure such as an RSU [16]. This type of system is also known as a *roadside to vehicle communication* (RVC) system, where the RSU could be a licensed stationary Base Station (BS) or Access Point (AP) operating at a specific and static location and frequency. The communication may be unidirectional (e.g.

broadcasting station) or bidirectional (e.g. vehicle communicating with RSU in P2P fashion).

In this system the synchronisation of the physical layer and medium access tasks are undertaken through an AP or a BS. Communication in V2R systems is made possible using cellular infrastructure networks, e.g. Global System for Mobile communication (GSM) and Universal Mobile Telecommunications System (UMTS), whichever applicable depends upon the region, and Wireless Local Area Network (WLAN) which is typically installed for inter/intra-vehicular communication. In certain scenarios, RVC systems are deployed in some areas to provide reliable broadband communication services, grant access to online resources, enable communication with other people and provide access to certain local services such as traffic or tourism-related information [16]. V2R systems can again be categorised as *Sparse* (SV2R) and *Ubiquitous V2R* (UV2R) communication systems [17]. SV2R systems provide information exchange and communication at hot spots e.g. at gas stations for price advertisement, or parking at the airport etc.

UV2R systems, on the other hand, need considerable investment in order to provide seamless, robust communication. Different RSUs might be connected via a wired backbone to form a wireless metropolitan area network (WMAN) and consequently a wireless wide area network (WWAN) [7, 27]. Figure 2.4 portrays an RVC system in which two RSUs are interconnected through a wired connection.

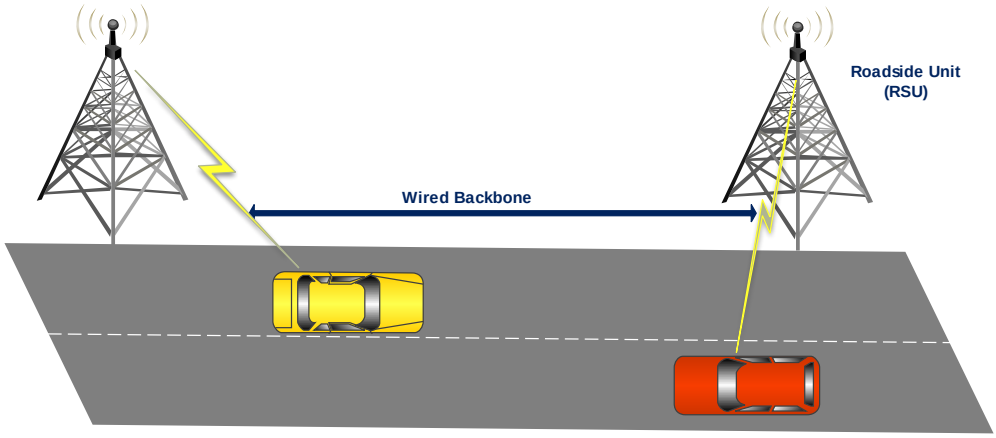


Figure 2.4 : Vehicle to Roadside communication system.

2.2.1.3 Hybrid vehicular communication (HVC) system

In *hybrid vehicular communication* (HVC) systems [17], nodes representing vehicles communicate with each other (V2V), with RSUs (V2R) and exploit both where it is applicable and/or practical. Each vehicle equipped with the proper instruments repeatedly broadcasts its status (location, speed, direction, etc.) to any nearby vehicles within its communication range and at the same time collects the same type of information from other vehicles [16]. Vehicles can inform other nearby vehicles of an accident or other potentially hazardous situations by deploying multi hop IVC (MIVC) or multiple RSUs in an HVC system (as shown in Figure 2.5) and thereby hopefully avoid any (further) collisions and consequently improve the safety of all parties. An HVC system can grant access to any online resources available to the vehicles not in the direct communication range of an RSU [17].

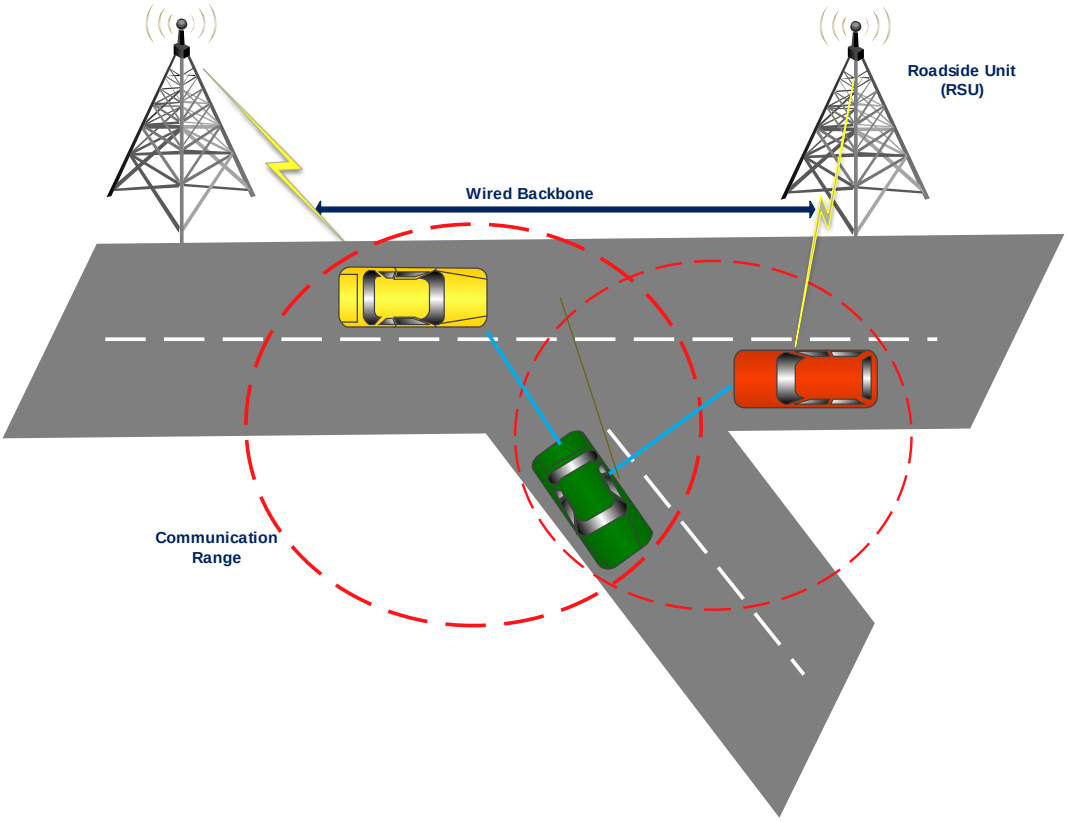


Figure 2.5 : Hybrid vehicular communication system.

2.2.2 Vehicular environments

Considering the fact that cost and performance trade-offs are the most important factors in setting up any communication, it must be admitted that different network configurations need to be explored based on different environments. The details of such an undertaking are discussed below.

A. Urban/city environment

In an urban or metropolitan environment, it is sensible to assume that wireless infrastructures are already deployed beforehand to provide universal connectivity in a cost-effective manner. The first of the major challenges in urban environments is the effect of shadowing caused by large buildings and other such objects, which significantly weakens the wireless channel. In addition, in a given traffic congestion scenario (e.g. a gridlock), the density of the vehicular nodes grows rapidly, bringing about a collision of transmissions resulting in poor system performance. In a city environment, multiple wireless technologies [17] may, and usually do, coexist in each other's proximity. These might include cellular services providing nearly seamless connectivity and Wi-Fi deployments which continue to become more and more widespread and common. Predicting traffic using these tools can lead to more efficient journey planning and an improved use of available resources [16].

B. Rural environment

In rural environments, IVC systems, augmented by limited infrastructures (covering hot spots or other areas of potential interest), might be considered as more economically efficient way of creating a communication network. In this kind of area the network is sparse and low in density because very few smart vehicles will be on the road [17].

C. Highway/motorway environment

The environment of highways or motorway is the site for the third scenario in which the benefits of vehicular networks can be demonstrated [36]. On a highway, the communication setup is usually hybrid. Installing RSUs at regular intervals might help support seamless communication. The highly mobile state of vehicles results in this environment posing various challenges such as scalability and the difference of the relative speed between vehicles and BSs. Utilising intelligent inter-vehicular communication, the vehicles can be allowed to select a faster and safer route in times of traffic congestion.

2.2.3 Applications supported by vehicular networks

The opportunity to design an intelligent vehicle safety system which is able to collect, process, and disseminate information, is guaranteed by adding a network interface, Global Positioning System (GPS) receiver and sensors. Inter-Vehicle Communication (IVC) puts forth a promising outlook in vehicular technology which must be enhanced by a variety of applications and software designed to simultaneously improve both safety and traffic situations [37, 38]. One can classify these applications into two basic categories: safety related and non-safety related applications [6, 16, 38, 39]. These two types can be further classified into several subcategories including applications which are aimed at traffic management, traffic coordination and assistance, tourism information and, finally, business and entertainment applications.

2.2.3.1 Safety related applications

Ensuring and increasing the safety of drivers, passengers and pedestrians by exchanging information on safety via IVC is the most important aspect of the intelligent transportation system [16]. This important task is undertaken by designing appropriate safety applications. From the 34 safety-related applications reported by the *Vehicle Safety Communication* (VSC), 19 are designed for RVC systems while the remaining 15 applications are intended to be used in IVC systems [40]. Several of these safety applications are described below:

Traffic signal violation: Since accidents caused by violating traffic light rules represent a significant part of traffic incidents, this application is quite useful. It functions via communication between intelligent vehicles and an RSU, as depicted in Figure 2.6. In this setup the RSU monitors vehicles approaching potentially hazardous intersections [6]. The traffic light status which will be constantly transmitted to all vehicles within communication range will be received by the on-board unit (OBU) in the vehicle, which then alerts the driver about the possible risks to which the violation of the traffic signal ahead might expose him/her [7, 39]. The RSU is also able to send information of any occurrence of traffic signal violation to the traffic management office or the police.

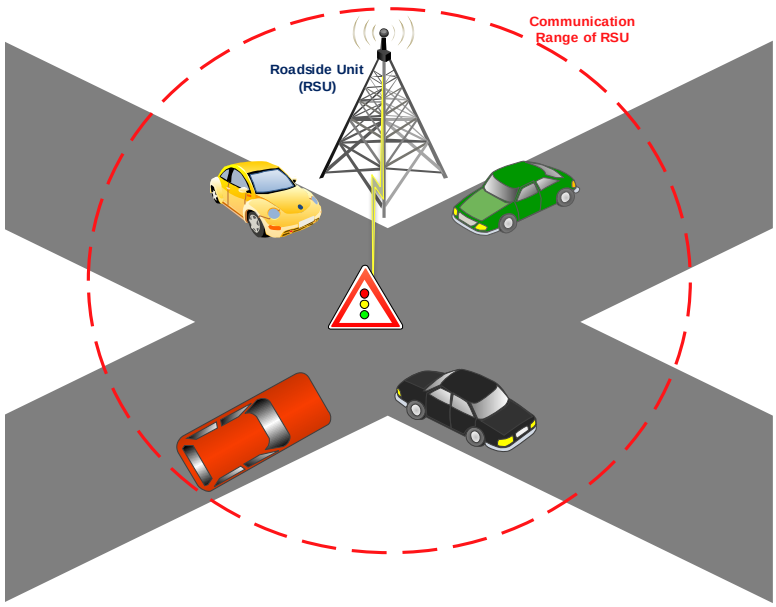


Figure 2.6 : Traffic signal warning.

Curve speed warning: On most roads, a sign indicating an upcoming curve is placed near a bend in the road in order to warn the drivers. The drivers will then estimate the necessary time and manoeuvre to reduce their vehicle's speed so as to safely pass the curve. However, considering that sometimes these estimates might not be accurate enough due to the human error factor, a curve speed warning application [7] is envisioned. This application requires an RSU placed near a potentially dangerous curve transmitting intermittent information to all approaching vehicles including the geometry of the curve, road conditions and any other incidents ahead. A typical example of curve speed warning can be seen in Figure 2.7.

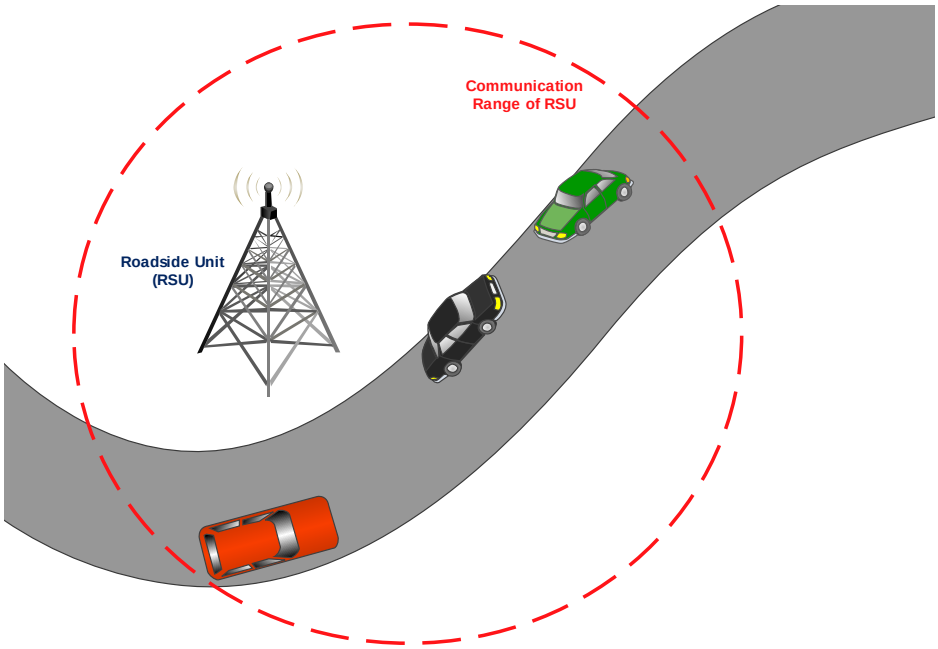


Figure 2.7 : Curve speed warning.

Emergency electronic brake lights: As a short term application [6, 7], emergency electronic brake light was expected to be deployed between a time frame of 2012 and 2016 but its deployment has not happened yet. Mainly used in conditions of reduced visibility (e.g. bad weather), this application is launched as soon as the brakes are applied by a vehicle: an emergency message is transmitted to all approaching and following vehicles, the OBU of these vehicles would warn the drivers and thereby ensure their safety as shown in Figure 2.8.

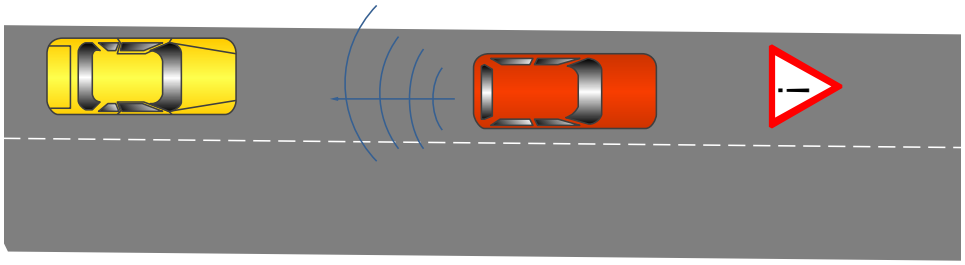


Figure 2.8: Emergency electronic brake warning.

Post-crash warning: In motorway/highway scenarios, a vehicle initiating an emergency deceleration might result in pile-ups due to its high speed and little reflection time of the drivers of the approaching vehicles [41]. A warning message sent by the affected vehicles to nearby approaching vehicles in time might reduce the level of damage. This post-crash warning system which works through broadcast in a multi-hop fashion might also be utilised in avoiding traffic congestion on the same road.

Cooperative driving assistance: In this type of safety applications [6, 41], the status of each vehicle is frequently sent to all nearby (in-range) vehicles. The vehicle's status is first obtained from GPS and OBU sensors. Each vehicle keeps a table of neighbouring vehicles' status. This information helps extend the knowledge of the driver regarding the road traffic conditions and potentially hazardous situations, which in turn results in safe and efficient driving.

2.2.3.2 Non-safety related applications

The FCC's allocation of seven 10 MHz channels allows other non-safety related applications to be deployed alongside of safety applications. The most

important aspect to consider when designing and deploying comfort applications is that they should not interfere with the safety applications which have higher priorities. One way to guarantee this is using a distinct physical channel. These applications can increase the quality and comfort of driving experience [17]. Weather information, maps, gas station and restaurant locations are examples of this type of application.

The information necessary for these applications can be provided with the help of traveller information application installed in vehicular systems [17]. In the context of business, commercial vehicles or enterprises can advertise their services using service channels [23]. A gas station can advertise the price of oil or any other services it provides by using an RSU. Furthermore, traffic management authorities can collect toll electronically with the help of RVC [7].

Of course one of the main functions of these applications would be that the passengers can entertain themselves by using online services such as instant messaging or playing online games with passengers in other vehicles [38]. The passengers can engage in various digital activities including checking their emails, downloading music or any other information using an Internet connection provided by a commercial RSU. The deployment of public safety and non-public safety applications requires exact evaluation as it depends on various interrelated parameters including cost, interoperability, infrastructure available and the density of vehicles on the roads.

2.3. Wireless Access in Vehicular Environments (WAVE)

The IEEE wireless access in vehicular environments (WAVE) working group develop and maintain the standardisation of 5.9 GHz DSRC technology [14, 42]. The IEEE 1609 working group undertakes the overall standardization of the communication stack from link to application layers [43, 44]. The WAVE standards stack includes the IEEE 802.11p standard which defines the physical and MAC layer features utilised in vehicular communication.

The WAVE protocol architecture offers support for the Transmission Control Protocol/Internet Protocol (TCP/IP) suite and the WAVE short message protocol (WSMP) in network and transport layers. The WSMP protocol has a very low delay to support optimised operation in the WAVE environment. This enables applications to control physical parameters such as transmitter power and the channel number used in transmitting messages [43].

2.4 Cloud and Fog Computing for Internet of Things

In a smart city vehicular environment, mobile users in the vehicles are main driving force towards the growth in traffic. From the data of 2013 to 2014, there has been a growth of 60% in mobile data traffic [45]. Researchers expected 8 times growth of traffic by 2020 [46]. Olariu et al [47] coined the idea of vehicular clouds by taking traditional VANET to the cloud. Vehicular cloud computing is a

new hybrid technology solution with impact on the Intelligent Transportation System (ITS) by using the resources of vehicles such as GPS, store internet and sharing information on the cloud [48]. A potential structure framework for vehicular cloud was proposed by [47]. European project TROPIC [46] brought the idea of bringing cloud capabilities closer to the mobile devices. A cloud to the ground support for smart things is called Fog computing, which is introduced by cisco [49]. Fog servers are positioned between cloud and smart devices. Fog provides low latency location awareness and improves quality of service (QoS) for streaming, real time application and supports heterogeneous devices and mobility [50].

Energy efficiency in vehicular domain (both vehicle to vehicle and vehicle to infrastructure) has not been thoroughly researched. Some effort has been devoted to optimising the location of fixed nodes like base stations (BSs) or roadside units (RSUs) to reduce network energy consumption [51-53]. While some savings are achieved through these techniques, further savings may be possible by reducing transmission energy consumption at a node by introducing sleep strategies during the inactivity periods [44, 54]. Introducing sleep cycles is an attractive solution for wireless networks as it does not require a complete overhaul of network devices, protocols or architecture. Such techniques have already been utilised for the line-cards in the routers [27, 55] where up to 79% reduction in energy consumption was achieved. Such major reduction may not be feasible in wireless and mobile network (e.g. cellular or vehicular) as they

are not intrinsically over-provisioned and the link quality is dependent upon the varying wireless channel, which makes it susceptible to degraded QoS. Nevertheless, a few research groups have proposed a number of sleep strategies to improve cellular networks energy efficiency [56, 57]. In [58], the authors proposed dynamic switching for a BS in low traffic conditions. However, fast switching may not be feasible to accommodate transient traffic behaviour because of the number of operations a large BS has to perform [57]. In a macro-micro cellular architecture [59], where small RSUs are used for offloading purpose, introducing various types of sleep cycles (random [60] and adaptive [61]) can be extremely effective due to the shorter resource activation time of an RSU. Therefore, such mechanisms are worth exploring in the context of a smart city vehicular environment. Using Fog-computing for vehicular content delivery, researchers are aiming to improve higher energy efficiency.

The authors in [62] proposed a sleep strategy which reduced energy consumption of the RSU by switching OFF only its transmitter part for a randomly distributed time duration when there is no request to serve. The RSU remained in sleep mode for a fixed time duration (randomly generated with a certain mean value) even if contents were waiting to be served. Upon waking up, the RSU served the arrived contents (if any) and switched back to sleep mode when the buffer became empty. This mechanism was called random sleep cycles. The same random sleep cycles were utilised in [30], which resulted in an unacceptable average content delay for audio-conferencing

applications [63]. This type of sleep cycles degrades the system performance, when a large amount of content waits to be served. Since energy saving through sleep cycles type-I was achieved at the expense of degraded QoS [30] and incurred wake-up overhead, associated with each sleep cycle [62], there is a need to improve QoS and reduce wake-up overhead while maximising energy savings. One way of achieving this is to put a limit (bound) on delay which enables one to find a maximum average sleep duration for that period (i.e. an hour in our case). Maximising average sleep duration for each hour in turn minimises wake-up overhead. To improve the performance of vehicular networks, a number of rate adaptation techniques has been utilised in the literature [64-67]. An exhaustive experimental evaluation of rate adaptation algorithms in real environments was presented in [64], followed by the development of a low-overhead rate adaptation algorithm which maximised the network throughput while minimising the bit error rate. The authors in [65] analysed the performance of rate adaptation techniques based on the concept of ‘coherence time’ using a channel emulator. Moreover, a rate selection policy was presented based on the speed and location of a vehicle [66]. Rate adaptation can also be utilised for energy efficiency as shown in [67]. Due to the environmental and economic factors besides having energy-efficient Fog computing for IoT Content Delivery (IoTCD), the use of renewable sources of energy such as wind energy is inevitable. The main challenge of using renewable energy in IoTCD lies in its judicial distribution for the best achievable QoS. Invariably this requires detailed traffic pattern. To the best of our

knowledge, transient performance analysis of Nano Servers (NSs) using sleep cycles, rate adaptation, and renewable energy in vehicular IoTCD has not been done before.

However, introducing rate adaptation in conjunction with sleep cycles redefines the performance from both QoS and energy perspectives. Further, using rate adaptation for energy savings in the Fog Servers in smart city not been done before.

Chapter 3: Load Adaptive Caching Points for a Content Distribution Network

3.1 Introduction

With ‘connectivity on the move’ being an integral part of our lives and the exponential growth in media rich content demand, substantial deployment in incumbent cellular third/fourth generation (3G/4G) wireless infrastructure has increased the overall network energy consumption many-folds. Therefore to cope with the ever-growing demand while keeping the energy consumption to a minimum, there is a need for energy efficient dedicated vehicular Content Distribution Networks (CDNs).

To decrease the overall network energy consumption, for the first time a switching off strategy was proposed in [68], which reduced the number of base stations. A similar strategy was employed on the CPs in a city vehicular CDN, where the number and locations of the CPs were optimised with vehicle mobility and hourly traffic [69, 70]. However, there would be transient variation in traffic within any period of time. Thus, there could be numerous occasions, where a CP has no request to serve. Thus, further reduction in network energy consumption is possible by efficiently scheduling CP transmissions, hence the

focus of this chapter. Reduction in transmission energy consumption at a CP can be obtained by switching OFF its transmitting circuitry for a random time duration when there is no request to serve. This process is called random sleep cycles [71]. Such reduction, though lesser in magnitude than that of the network energy savings, enhances the overall energy savings. However this is generally achieved at the expense of degraded QoS [71]. Therefore, it is worthwhile to study the trade-off between QoS and energy savings through transient analysis of the city vehicular CPs instead of day wise or hourly steady state analysis. The QoS can substantially be improved through the introduction of load adaptation, which enables the CP to operate at variable data rates depending upon the load. This leverages maximum energy savings with acceptable QoS. The introduction of our load adaptive technique and random sleep cycles redefines the performance of CPs in a city vehicular CDN. Therefore, our contributions in this chapter are the following:

A detailed transient load analysis has been carried out at both primary and secondary CPs in terms of piece request arrival rate and number of simultaneous connections.

Based on [72], the load adaptive service discipline for the proposed CP is implemented through det-Neg distribution.

A be-spoke simulator[1] has been developed in Java, which has mobility module and communication module simulating request and service to obtain performance results, in terms of energy savings, average queue size, average piece delay, and average content delay utilising typical vehicular traffic profiles gathered at the city of Saskatoon, Canada [73].

This chapter is organised as follows: Section 3.2 a brief discussion on related work. An overview of the studied scenario is presented in Section 3.3. In Section 3.4, a simulator developed for the CPs with multiple random sleep cycles and load adaptation. Section 3.5 the performance of the CPs with different sleep durations evaluated. Finally, the chapter concludes in Section 3.6.

3.2 Related Work

Content caching[2] and distribution are attracting considerable attention in vehicular networks [44, 52-55]. The authors in [52] introduce a caching scheme that uses relays to offer VoD playback service with a lower delay. In [53], the authors proposed a content distribution scheme for VANETs with the aim of achieving low latency content distribution in a dense vehicular highway network. RSUs were used to distribute the content to vehicles in a unidirectional highway

[1] Be-spoke simulator is specifically designed to meet our requirement.

[2] Content caching is a type of performance optimisation in which files are loaded from nearby servers instead of a single origin server. When enabled, user save time by loading content from a local server of a potentially distant one.

network, with the help of V2V communication to achieve better connectivity and lower latency. The authors in [44] proposed a push-based popular content distribution (PCD) scheme in which RSUs proactively broadcast popular content to the vehicles in an area of interest. The vehicles that were within the range of each other form a VANET to distribute the received content among them. A symbol-level network coding (SLNC) technique was used to address the network fragmentation problem and to reduce the overheads while maintaining the desired performance. In [54], the authors proposed a Cooperative Content Distribution System for Vehicles (CCDSV) which utilised a set of access points (APs) that cooperated in disseminating content to vehicles. The proposed scheme addressed a number of issues such as the quality of mobility prediction and the lack of APs resources. A Vehicular Content Distribution (VCD) is proposed in [55], which enables high bandwidth content distribution in vehicular environments. In VCD, a link between a vehicle and an AP is opportunistically established. In order to efficiently utilise such links, the content is proactively pushed to the APs that vehicles are likely to visit in the near future (which are predicted via a dedicated algorithm). Hence, the entire wireless capacity of the AP is utilised efficiently rather than experiencing the bottlenecks of the traditional Internet connectivity.

There have been a number of research efforts in the recent past to make wireless networks energy efficient. While one of the views is to optimise RF output power [27, 56] of wireless nodes, others [58] found of little use as the

power consumption of the circuitry can be much higher than the transmitter output power. Collectively, maximum energy savings in a wireless network can be obtained with reduced transmission power and optimised operational power. Considerable effort has been devoted to optimising the location of fixed nodes like base stations (BSs) or roadside units (RSUs) to reduce network energy consumption [57, 59]. While the bulk of the savings was achieved through these techniques, further savings may be possible by reducing the transmission energy consumption at a node by introducing sleep strategies during the inactivity periods [60, 61]. Introducing sleep is an attractive solution for wireless networks as it does not require a complete overhaul of network devices, protocols or architecture as such techniques have already been utilised for the line-cards in the routers [65, 66] where up to 79% reduction in energy consumption was achieved. Such major reduction may not be feasible in wireless and mobile network (e.g. cellular or vehicular) as they are not intrinsically over-provisioned and the link quality dependent upon the varying wireless channel, which makes it susceptible to degraded QoS. Recent research on energy savings encompasses various methods, where the principal objective is to maintain quality of service (QoS) for communication networks [30, 70, 74]. Nevertheless, a few research groups have proposed a number of sleep strategies to make cellular network energy efficient [67, 75, 76]. In [75], the authors proposed dynamic switching for a BS in low traffic conditions. However, fast switching may not be feasible to accommodate transient traffic behaviour because of the number of operations a large BS has to perform [76].

In a macro-micro cellular architecture [77], where small RSUs are used for offloading purposes, introducing sleep cycles (random [71]) can be extremely effective due to the shorter resource activation time of an RSU. Therefore, such mechanisms are worth exploring in the context of a vehicular CDN.

To improve the performance of vehicular networks, a number of rate adaptation techniques have been utilised in the literature [78-81]. An exhaustive experimental evaluation of rate adaptation algorithms in real environments was presented in [78], followed by the development of a low-overhead rate adaptation algorithm which maximised the network throughput while minimising the bit error rate. A rate selection policy was presented based on the speed and location of a vehicle [80]. Rate adaptation can also be utilised for energy efficiency as shown in [81]. However, introducing load dependency in conjunction with sleep cycles redefines the performance from both QoS and energy perspectives in a city vehicular environment.

3.3 The Studied Scenario

A vehicular CDN for a city environment is considered (shown in Figure 3.1), where the number and locations of the CPs are already optimised to reduce network energy consumption by a previous PhD student [82]. The traffic is generated by the moving vehicles in the form of requests for pieces of a video content stored in the CPs. Vehicular content delivery is assumed to be through an idealistic (contention and collision free) MAC protocol between the vehicles

and the CP. This assumption is not far-fetched as with realistic channels, the traffic arriving at the CPs is reduced only by a fraction due to loss, hence not the focus of this chapter. A CP provides coverage of 200 meters (diameter) and the maximum speed of a vehicle in a typical city is considered to be 46 km/h. Therefore, a vehicle stays in the range of a CP for a minimum of 16 seconds. The impact of handoffs is not considered as the content file is divided into smaller pieces in such a way that typical vehicular mobility and download rates result in vehicles being within the range of a CP until a piece of size 6 MB is completely downloaded (even when vehicles travel at the maximum legal speed in the city). Hence the content of size 200 MB [83] is divided into 33 equal pieces. The mean inter arrival time of piece requests is considered as 16 seconds and is Negative exponential distributed.

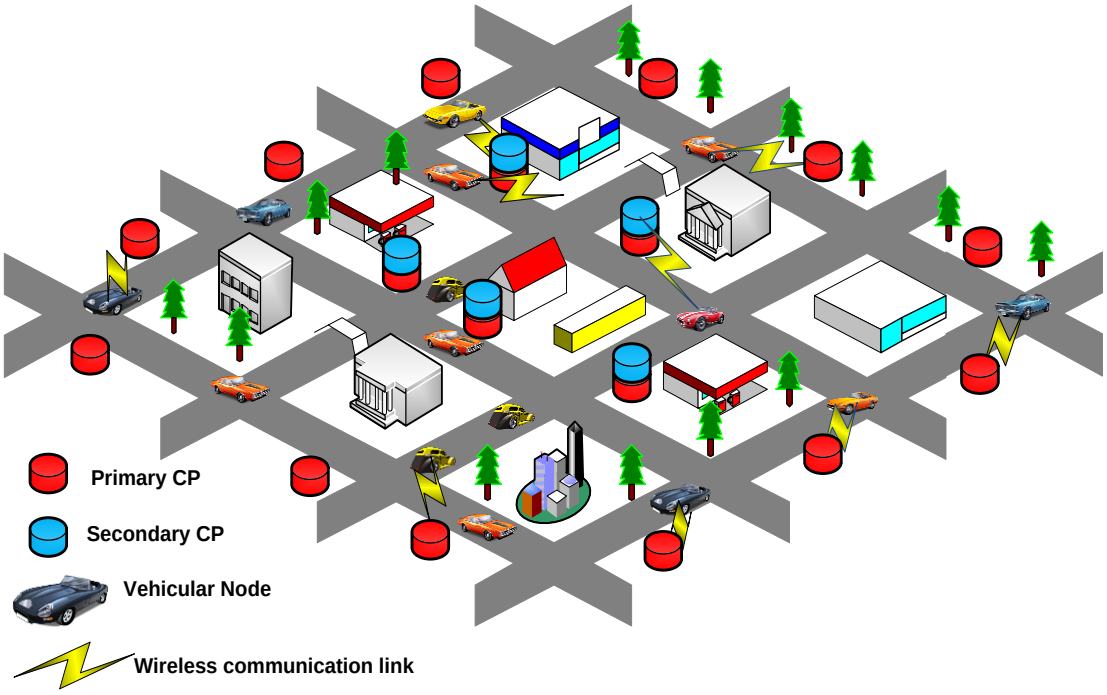


Figure 3.1: City vehicular scenario.

It is assumed that as soon as the number of simultaneous connections (piece downloads) exceeds 10 at a CP, a new CP is switched on to cater for the remaining data traffic. The former CP is termed as primary while the latter as secondary. The primary and secondary CPs are co-located so that the secondary CPs can immediately be activated on demand to accommodate excess traffic. The arrival rate of piece request (shown in Figure 3.2) are the inputs to the proposed simulation model for the primary and secondary CPs. The total number of piece requests within an hour results in the arrival rate and service rate (shown in Figure 3.2 and Figure 3.6, respectively) at each CP whereas the number of simultaneous connections (each operating at 3 Mb/s) reflects the variable data rate within that hour (see Figure 3.6). To minimise

transmission energy consumption at a CP, multiple random sleep cycles operate in the following way. When the CP is idle (i.e. its buffer is empty) it switches to sleep mode for a random amount of time with a certain mean duration in order to save transmission energy. In order to implement energy savings through sleep cycles, the distribution of sleep durations is assumed to be negative exponential considering Poisson distributed arrival process of piece request and negative exponential distributed piece service discipline. The transmitter power consumption of a CP is determined as $P_{MAX} - P_{MIN} = 7.8\,W$ [84]. Upon waking up, if there are requests waiting in the buffer to be served, the CP serves them, otherwise, the CP switches to sleep mode again. Figure 3.3 shows a queuing model of a CP with load adaptation. All system parameters are summarised in Table 3.1.

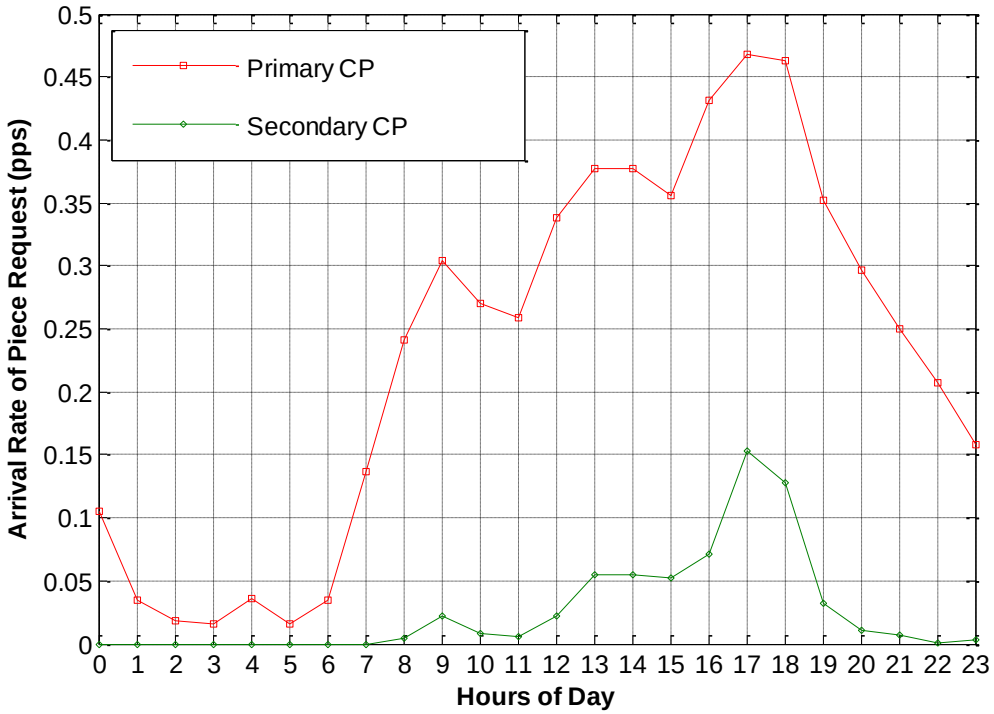


Figure 3.2 : Arrival rate of piece requests at primary and secondary CPs for varying hours of the day.

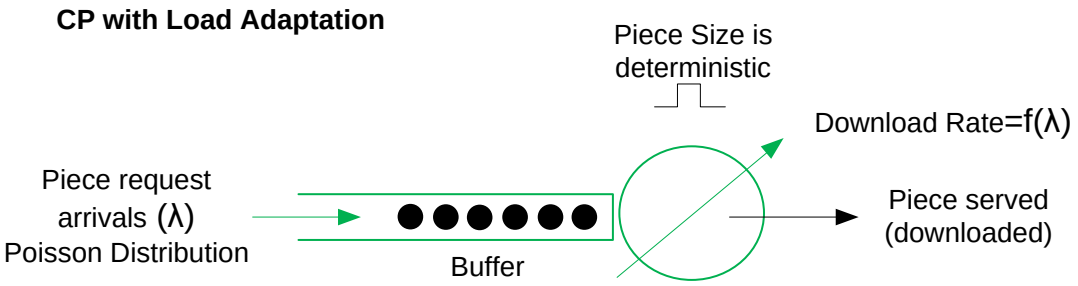


Figure 3.3: Queues with load adaptive data rate and fixed piece size.

Table 3.1: System parameters.

Content size	$200\ MByte$
load adaptive data rate (dynamically allocated)	$3\ to\ 30\ Mb/s$
Piece size	$6\ MByte$
Transmit power of CP	$7.86\ W$ [84]
Energy for wake-up overhead (E_{wo})	$0.0175\ J$ [84]
Number of pieces per content	33
Transmission range (radius)	$100\ m$
Mean inter arrival time of piece request	$16\ s$

3.4 Load Adaptive CPs Simulation

The load adaptive CP works on the concept of statistical multiplexing, where parallel simultaneous connections are multiplexed into a single serial pipeline. It is an established theory that a server with Ck capacity is superior to the k parallel servers, each with capacity C in terms of latency [84]. Since the piece size is fixed (deterministic), where each piece is ensured to receive an average download rate between 3 Mb/s and 30 Mb/s (corresponding to 1 to 10 simultaneous connections). The download rate at CP works depends on the number of requests queuing in its buffer, therefore the service duration of a piece can be represented as a det-Neg distribution [72]. The flowchart of load adaptive CP is presented in figure 3.4.

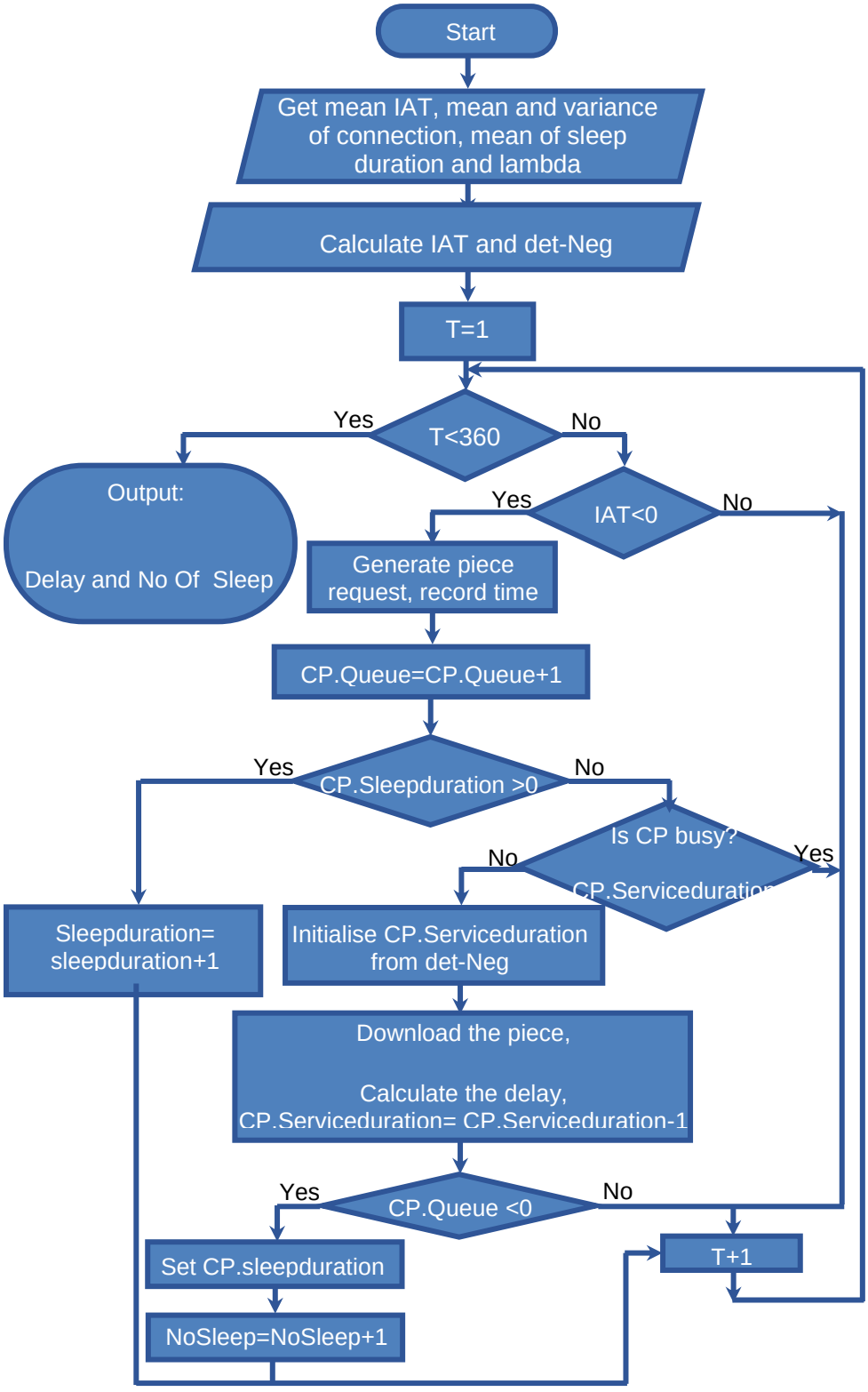


Figure 3.4 : Delay and number of sleep for a Load Adaptive CP.

3.5 Performance Evaluation

The performance of the network has been evaluated in terms of average piece delay, average content delay and transmission energy savings with respect to load adaptive data rate and average sleep cycle durations of 100 ms and 1 s. The load adaptive data rate reflects varying vehicular density throughout the day. In traditional access networks, packet size can be fixed/variable but the data rate is generally fixed. However, in the present case, pieces with fixed size are served with a load adaptive data rate, which statistically multiplexes (aggregates) up to 10 simultaneous connections. Hence, the proposed CP in this chapter can serve at data rates that vary between 3 Mb/s and 30 Mb/s depending on the number of simultaneous connections rather than serving with a fixed data rate of 3 Mb/s/user. Note that the buffer is assumed to be infinite as the cache point with storage can virtually hold infinite number of piece requests.

3.5.1 Average Piece Delay

Figure 3.5 shows the average piece delay served by the primary and secondary CP throughout the entire day. The average piece delay does not follow the trend of the load (Figure 3.2) as the data rate here is load adaptive. Unlike the usual case (with traditional CPs) where the increase in average piece delay is only governed by the offered load, the average piece delay in the proposed CP is primarily dependent on adaptive data rate (Figure 3.6), sleep cycles and their transient variations. Therefore, as the main goal is to increase transmission rate

at higher traffic to maintain constant delay. Figure 3.5 shows success in this endeavour. The number of times a CP operates sleep cycles is inversely dependent on load. Thus the waiting delay for a piece due to sleep cycles does not increase at high load. In contrast, the serving duration (download time) decreases at higher load due to the adaptive data rate. This decreases the overall average piece delay at higher load. It is to be noted that the load adaptive CP never operates at its maximum value throughout an hour. Thus, the mean data rate never reaches 30 Mb/s.

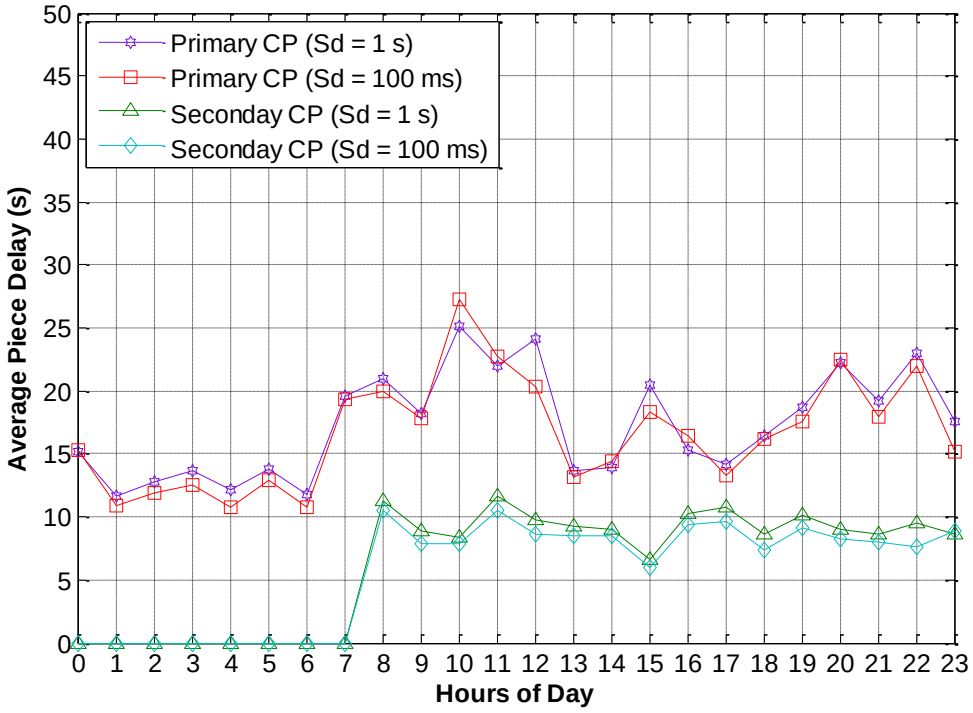


Figure 3.5 : Average piece delay for varying hours of the day.

Figure 3.6 shows that the mean service rate follows the traffic profile during the day (Figure 3.2).

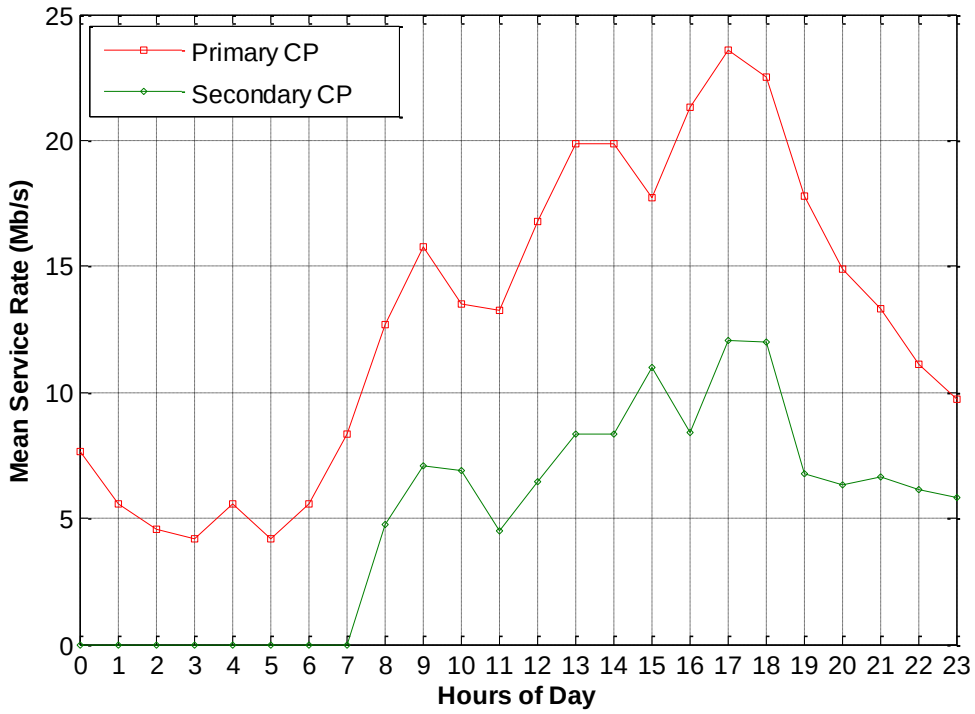


Figure 3.6 : Average service rate (Mb/s) for varying hours of the day.

3.5.2 Average Content Delay

Figure 3.7 shows the average content delay at both CPs for each hour of the day. The figure shows that the average content delay has a similar behaviour as that of the average piece delay. This is because the content consists of 33 pieces, where the average content delay is computed as the sum of the average piece delay of 33 pieces for estimation purpose (without considering mobility).. The download variation for pieces between a CP to the next CP is not

considered. Moreover, since the mean content delay depends on the average piece delay, the former exhibits similar behaviour.

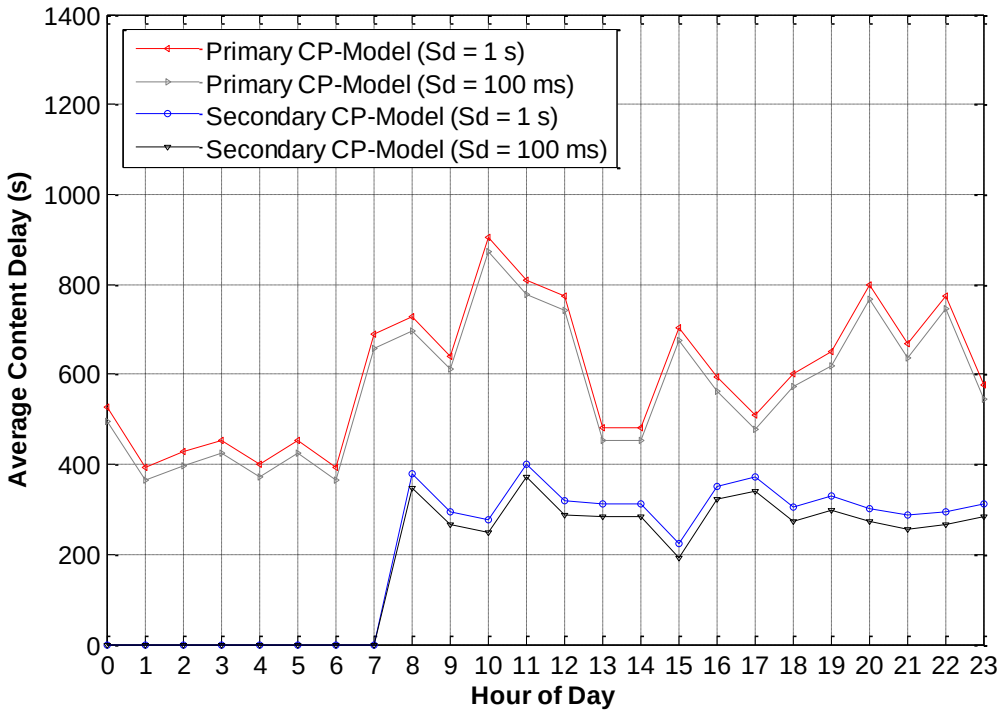


Figure 3.7 : Average content delay (estimated) for varying hours of the day.

3.5.3 Energy Savings

Figure 3.8 illustrates the energy savings achieved by both primary and secondary CPs throughout the entire day by operating multiple random sleep cycles with different mean durations. The figure shows that the energy savings achieved at the early hours of the day is higher due to the low vehicular load at the early hours that leads to a lower number of generated requests. This allows the CP to stay in the sleep mode for longer aggregated sleep duration in those

hours and hence achieving higher energy savings. For example, the energy savings achieved by the primary CP at 03:00 hr is 23 kJ compared to only 4 kJ achieved at 17:00 hr. Moreover, the figure also shows that the secondary CP achieves higher energy savings as compared to the primary CP. This is because the load on the primary CP is much higher than that on the secondary CP (please see figure 3.2). The energy savings achieved by the secondary CP between 00:00 hr and 07:00 hr is approximately 28 kJ compared to the 11 kJ achieved at 17:00 hr. The figure also demonstrates that the sleep cycle duration has insignificant effect on the results of both CPs. This is mainly because of the relatively large piece size (i.e. 6 MByte), service time and very low total wakeup overhead, compared to the traditional packet-switched networks. Note that the total wakeup overhead is very low due to the smaller number of times the CP can switch to sleep mode. Since the CP operates sleep cycles only when there is no piece request waiting in the buffer, the probability of a request arriving at a CP during the service is very high.

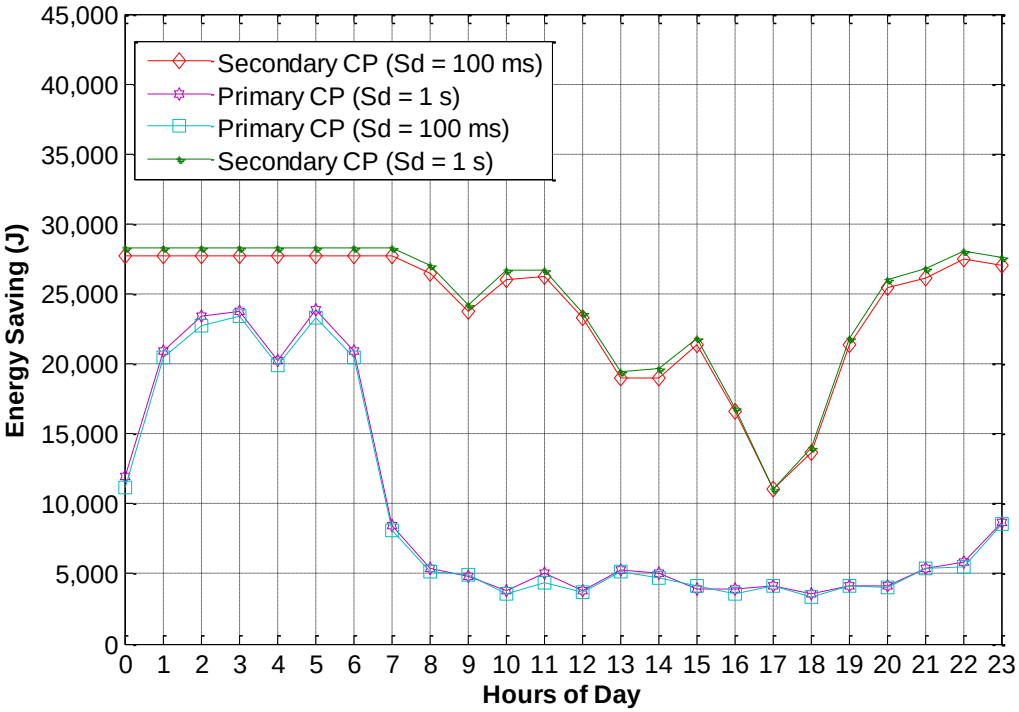


Figure 3.8 : Variations of energy savings for different hours of the day.

3.6 Conclusions

In this chapter, energy efficient load adaptive CPs for vehicular CDN are proposed. The load adaptive data rate obtained by statistical multiplexing of parallel piece downloads was implemented through det-Neg distribution. Such a CP was simulated as an M/det-Neg/1/∞ queue with queue length dependent vacations, where typical city vehicular traffic profiles were utilised. M/det-Neg/1/∞ queue is a special type of queue, where the arrival request is Poisson, service request follows a generic distribution, which varies between a pure deterministic behaviour and pure random (i.e. negative exponential behaviour).

It has a single server and considered to have very large buffer, which is theoretically infinity as it only holds requests.

The average piece delay at the proposed CP was dependent on adaptive data rate, sleep cycles and their transient variations. Since, the number of times a CP operates sleep cycles was inversely dependent on load, the waiting delay for a piece due to sleep cycles did not increase at high load. In contrast, the serving duration (download time) decreases at higher load due to the adaptive data rate. This decreases the overall average piece delay at higher load. The advantage of the proposed load adaptation (statistical multiplexing) technique at a CP operating sleep cycles is evident in terms of lower average piece delay and higher energy savings compared to that of a traditional CP. The performance results revealed that the proposed CP saved up to 84% transmission energy during off-peak hours and 33% during the whole day while fulfilling content demand in a city vehicular environment. Moreover, secondary CPs, due to their lower load, have achieved higher energy savings during their operating span compared to that achieved by the primary CPs. Furthermore, the results revealed that the mean sleep cycle duration has insignificant effect on the energy savings of both CPs due to the large piece size (i.e. 6 MB), long service time and low wake-up overhead.

In the next chapter, the performance of the caching points with renewable energy is studied for the content distribution network.

Chapter 4: Standalone Green Cache Points for Vehicular Content Distribution Networks

4.1 Introduction

Much attention has been drawn to Content Distribution Networks (CDNs), which has to do with the growing interest in media-rich files. Among the services that CDNs offer are Internet Protocol and Catch-up TV (IPTV and CuTV) and Video-on-Demand (VoD) [85, 86]. Thus, large bandwidth for the transmission and capacity for the storage of multimedia files are only two requirements. Content caching generally enables communication networks to lower end-to-end (e2e) delay and network congestion and also improve download rate. The problem of congestion can be settled by directing a BS's traffic to nearby Cache Points (CPs) with higher capacities an approach that significantly lowers the downloading time as well. Additionally, drawing content closer to end users considerably reduces energy consumption since road-side caches reduce journey up and down the network to fetch content and function at locations with weaker BS signals [87]. The above advantages warrant more attention in vehicular networks. Based on [88], by 2020 above 91% of the total IP traffic in the world will be a form of video, with a 33% annual increase. The increasing

interest shown by vehicular users to download multimedia files alongside the mentioned growing figures will even further exacerbate the problem of desirable quality of service (QoS) if energy saving is also to be considered. Moreover, The information and communications technology (ICT), according to [89], yields 2%-2.5% of the world carbon emissions, a number which is predicted to significantly grow in the next few years. Concerns over carbon footprints, coupled with the rapid reduction in fuel resources, have led to serious re-evaluations of ICT systems for both energy consumption and quality of service (QoS). To achieve reduce the carbon foot print of vehicular networks we propose the use of mini stand-alone wind turbines alongside RSUs, which eliminates grid power consumption. Furthermore, for improving QoS and to avoid dropping due to insufficient available renewable energy, we propose energy adaptive CPs where the number of simultaneous downloads is based on the amount of available renewable energy. To reduce network energy consumption, we previously optimised the number and location of CPs to support content delivery at high data rates, then for trade-off between having all renewable energy consumption and serving all traffic demand, the needed number of CPs are turned on. Therefore, our contributions in this chapter are the following:

We developed (i) Mixed Integer Linear Programming optimisation model (MILP) with non-renewable energy sources for non-adaptive CPs, (ii) MILP optimisation model with only renewable energy sources for non-adaptive CPs, (iii) MILP

optimisation model with renewable energy sources for adaptive CPs and (iv), Heuristic for all the proposed MILP models, to validate, compare and study the performance results in detail.

4.2 Vehicular City Model

To analyse the performance of the cache points, a simplistic model city is adopted in this chapter as shown in Figure 4.1. The model consists of 16 junctions and 48 roads, where each junction has 4 exits [4]. We have used typical traffic statistics of a UK city, on which statistical analysis has been carried out in order to capture traffic patterns throughout an office day. Our study includes data from October 15, 2009, and limits itself to a 9 km^2 coverage area, which means a maximum number of 500 vehicles per hour. The hourly vehicular flow is shown in Figure 4.2, where the peak traffic is observed in business and afternoon (16:00-18:00) hours.

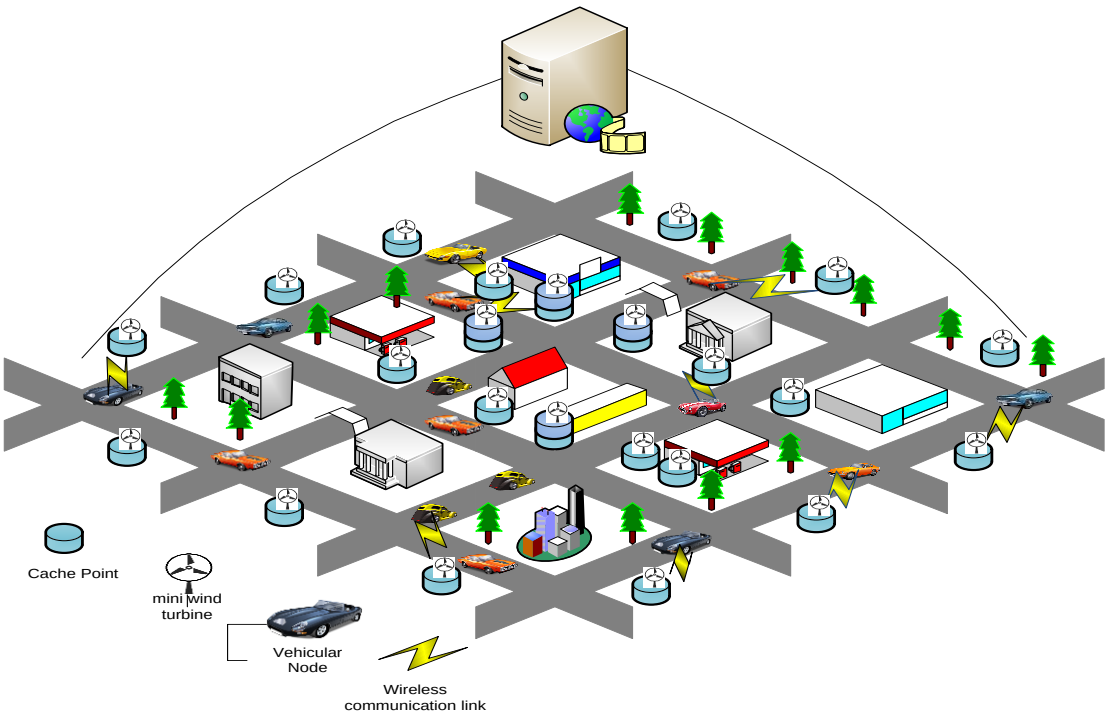


Figure 4.1 : City vehicular network.

4.3 Wi-Fi based Content Distribution Network

We propose a set of CPs with media-rich contents closer to the vehicles for decreasing the obstruction in the communication path between the server and the vehicles. Due to the limited storage capacity of the CPs, only the most popular media contents are stored. Instances of such popular content include advertisements in a city, movie trailers, videos on free parking spaces and BBC iPlayer and YouTube videos. Higher download rates and better energy efficiency are expected to be achieved by such a setup employed for coverage and CPs are continuously activated and deactivated to save energy [55].

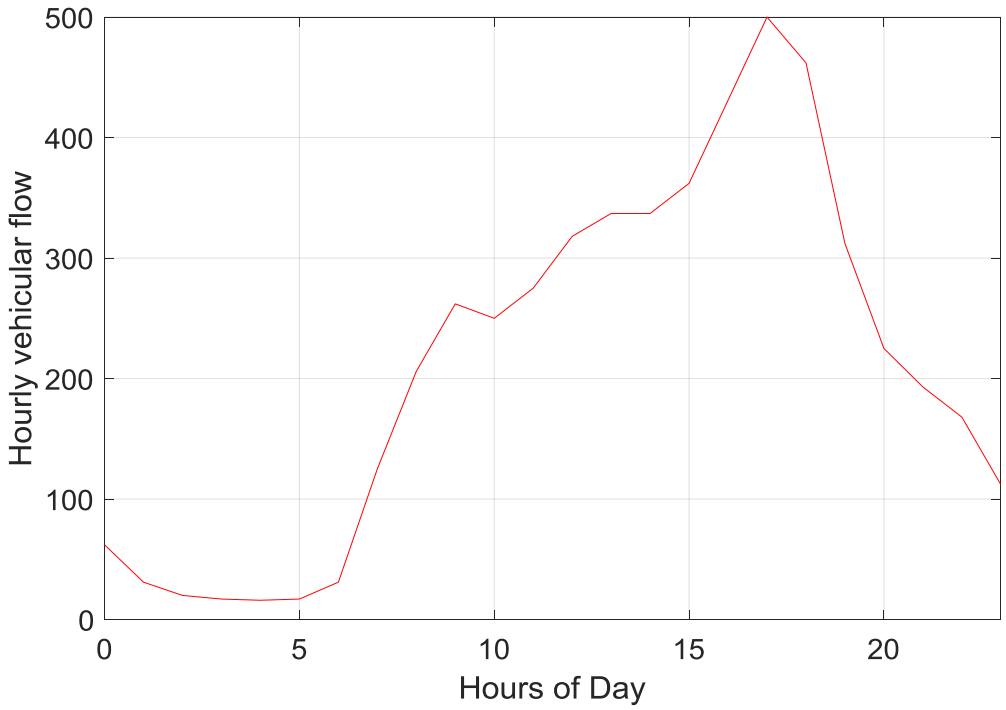


Figure 4.2 : Vehicular flow for varying hours of the day.

Considering the city set up, we identify 192 candidate sites (CSs) throughout the city with each road having four CSs. The candidate sites are the potential places where the CPs can be installed. These 192 CSs are uniformly distributed in the city, where each road has four CSs. Note that on each pair of roads (i.e. one road in each direction), a total of 8 CSs are available (i.e. four in each direction). This provides the basis of optimisation of the number and location of such CPs. Each CP is implemented with an IEEE 802.11p radio transceiver unit having data rates varying between 3 Mb/s and 18 Mb/s depending on the range of communication and the reliability of the channel [6]. According to [6], the typical range of communication in vehicular networks can reach up to 300 meters (radius). However, the current study assumes a range of 100 meters to

account for harsh city propagation environment [90]. This leads to higher number of CPs. Each CP is further equipped with a storage disk for a maximum of 50 content files, a microcontroller acting as an interface with application coordinates requests and orders files, and a MAC algorithm unit is used (e.g. Vehicular Content Download Algorithm VCDA [2]).

A content file is downloaded from multiple rather than single CPs during the entire journey of a vehicle. Each file is divided into a number of smaller pieces to avoid interruption during handover, where the complete file download is finished when a vehicle has downloaded all the pieces from different CPs. For instance, a CP with a 100 meters transmission range (200 meters in diameter) and a vehicle with a maximum speed of 13 m/s would lead to a case where the vehicle remains within the CP range for at least 16 seconds so a download rate of 3 Mb/s, a piece with a size of 6 MB can be downloaded (during the 16 seconds), where a typical YouTube 200 MB video consists of 33 pieces. The traffic is generated by the moving vehicles in the form of requests for pieces of a video content. The video content is replicated in all CPs. The maximum number of simultaneous connections (piece downloads) is computed from the available wind energy at each CP, which varies with time. As the MILP model is unable to handle the movement of vehicles, a set of Traffic Points (TPs) that represent traffic which account for the content/data traffic requested by nearby vehicles is assumed. Hence, a set of fixed traffic-generating points in the city space is seen by the MILP model. Each bi-directional road has one TP, i.e. a

total of 24 TPs are present in the city vehicular network. Only 4 out of the 8 candidate sites in each road are reachable from a TP. In a practical city the concept of TP doesn't exist so vehicles can use all the 8 candidate sites. The introduction of a single TP per road segment further reduces the MILP execution time.

4.4 Greening City Vehicular Networks

Due to ecological and economic factors, depleting fuel resources and global warming, introducing energy efficiency in information communication technology (ICT) sector has become a necessity.

4.4.1 Wind Power Profile

According to the U.S. Department of Energy (DOE), wind energy is the world's fastest-growing non-conventional energy. The UK is one of the best locations for wind power in the world [91]. At the beginning of 2014, the installed capacity of wind power in the United Kingdom was 8,445 megawatts (MW), with 362 operational wind farms and 4,158 wind turbines [92]. The United Kingdom is considered to be the world's sixth largest producer of wind power [93].

We did not consider solar energy as solar energy is typically not available for approximately half the day (for example in many locations in the UK, solar energy is not available for between 8 hours and 16 hours in the day depending on season [94]). This means that the batteries needed to support continuous

operation with solar energy are larger in size compared to those needed when wind energy is used [95]. The study of our techniques when solar energy is used is an interesting research topic that can be considered in future. The goal for this study was to examine the potential of wind in providing acceptable QoS. The first step is calculating the harvested energy from the wind in typical cities. The wind speed values for this analysis have been obtained from the UK air information resource (AIR) database provided by the Department for Environment, Food and Rural Affairs [96]. The data is comprised of hourly wind speeds for the whole of 2014 measured at one of their monitoring sites in Leeds City Centre, UK. In view of the wind variation and as we are considering a day's vehicular traffic profile, it was necessary for us to derive the expected value of generated wind power for each hour. Wind power expected value is considered for each hour however it will be days with insufficient wind generated power.

4.4.1.1 Mathematical Model

Wind turbines work by converting the kinetic energy in the wind first into rotational kinetic Energy in the turbine and then electrical energy that can be distributed. The wind power harnessed depends on the wind speed and the swept area of the turbine [97]. Table 4.1 shows the definition of various variables and parameters used in this model:

Table 4.1 : List of Symbols.

Parameters	Description
E	Kinetic Energy (J)
ρ	Air density (kg/m^3)
m	Air mass (kg)
A	Swept Area (m^2)
v	Wind Speed (m/s)
C_P	Power Coefficient
$\frac{dm}{dt}$	Mass flow rate (kg/s)
x	distance (m)
$\frac{dE}{dt}$	Energy Flow Rate (J/s)

The kinetic energy of a mass in motions is:

$$E = \frac{1}{2}mv^2 \tag{4.1}$$

The power in the wind is given by the rate of change of energy [95]:

$$P = \frac{dE}{dt} = \frac{1}{2}v^2\frac{dm}{dt}. \tag{4.2}$$

As mass flow rate is given by

$$\frac{dm}{dt} = \rho A\frac{dx}{dt} \tag{4.3}$$

and the rate of change of distance is given by

$$\frac{dx}{dt} = v. \tag{4.4}$$

We get

$$\frac{dm}{dt} = \rho A v. \tag{4.5}$$

Hence, from Equation (4.2), the power can be defined as

$$P = \frac{1}{2}\rho A v^3. \tag{4.6}$$

Wind turbines cannot convert all the kinetic energy of the wind into mechanical energy turning a rotor. C_P is the coefficient of performance of the wind turbine which accounts for the decrease in the actual power harnessed from the wind due to several factors such as, friction and equipment losses. Therefore, the power generated becomes

$$P_W = \frac{1}{2}C_P\rho A v^3. \tag{4.7}$$

The relationship between generated energy and wind speed is nonlinear. This means we cannot simply use the average wind speed to determine the average power generated. Therefore rewriting the generated power in terms of average values [96]

$$P_{avg} = (\frac{1}{2}C_P\rho A v^3)_{avg} = \frac{1}{2}C_P\rho A v_{avg}^3. \tag{4.8}$$

In probabilistic terms the expected value of v^3 is

$$v_{avg}^3 = \sum_i [v_i^3 P(v = v_i)] \quad (4.9)$$

One year wind data has been obtained to compute the hourly mean and variance values. We used these to compute the parameters of the stochastic process for a day then generated random numbers many times (i.e. many realisations) to obtain the results. Figure 4.3 shows the expected value of wind power of our case study city (Leeds). The results are based on mean wind power and it should be noted that there will be days when there may not be enough wind generated power. The wind power calculation system parameters are presented in Table 4.2.

Table 4.2: System parameters.

Parameters	**Values**
Propeller length in diameter (D_{jt})	$0.5\,m$ [96]
Swept area (A)	$0.2\ m^2$
coefficient of performance (C_P)	0.45 [96]
Air density (ρ) at 15^0 C	$1.225\ \ kg/m^3$ [94]

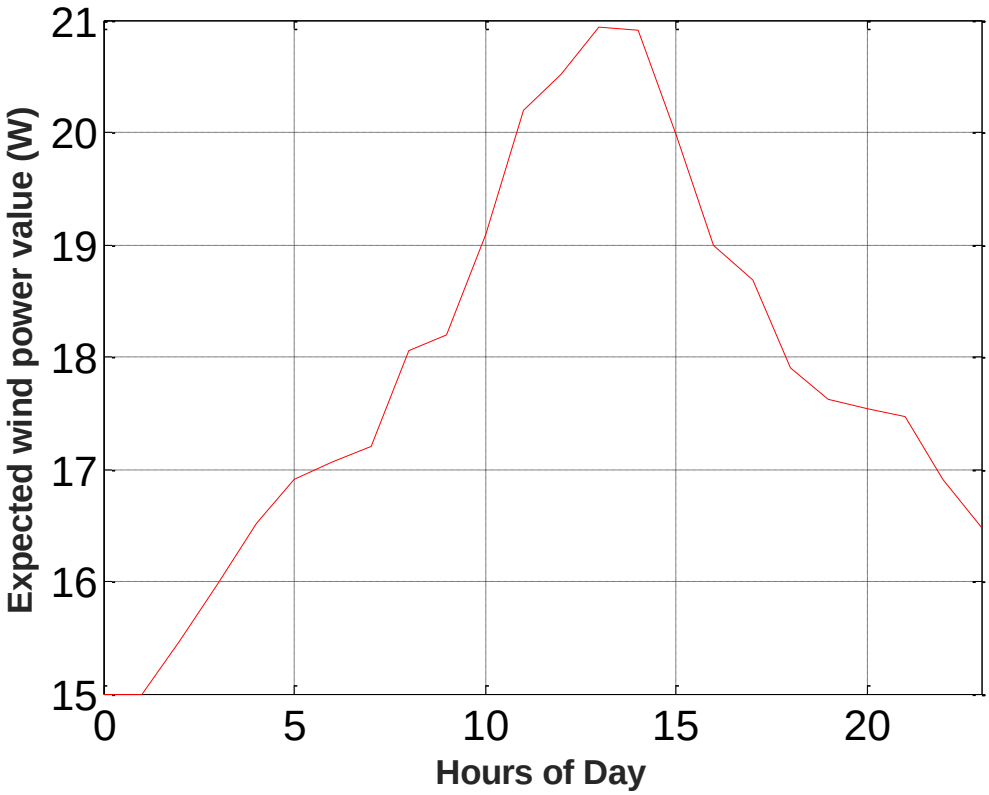


Figure 4.3 : Expected (Mean) wind power value in a typical UK site[3].

4.5 Vehicular traffic profile

The vehicular traffic is obtained from the city of Leeds, UK [99] by averaging and scaling the weekdays, Saturdays and Sundays traffic. The number of vehicles in our city area varies within 16 and 500. Upon reaching a junction, a vehicle chooses one of the four directions with certain probabilities. Out of 4 directions at each junction, two directions lead towards the city centre (in bound) and the other

[3] The results are based on mean wind power, it should be noted that there are days and some hours of a few days, when there may not be enough wind generated power, in which case either the RSU does not operate or it operates with a connected battery.

two lead out of the city (out bound). The probability of choosing a direction towards the city centre is variable depending upon the time of the day (0.1 to 0.9). Considering typical business traffic, we assume the direction probabilities as follows. Between 00:00 hour and 06:00 hour, 50% vehicles are inbound and 50% vehicles are outbound. During 07:00 hour to 09:00 hour, 90% vehicles are inbound (peak hours) and 10% vehicles are outbound. Between 10:00 hour and 17:00 hour, 70% are inbound and 30% are outbound. Between 18:00 hour and 21:00 hour, 10% are inbound and 90% are outbound (opposite of peak hours). Finally, from 22:00 hour to 23:00 hour, 50% are inbound and 50% are outbound. Each vehicle may generate request, which follows Poisson distribution with a mean inter arrival time (τ) of 2000 seconds. Based on the discussed detail, a bespoke JAVA-based vehicular mobility simulator for content delivery is developed and presented as flowchart in Figure 4.5. The simulator is run for 24 hours to generate traffic at various locations within the considered area for each hour of the day. Figure 4.4 shows hourly vehicular flow where business hours observe moderate traffic and late afternoon hours (16:00-18:00 hours) represent peak vehicular traffic. In addition, the resulting expected traffic demand value per second is also shown in Figure 4.4, which varies between 15 Mbps and 670 Mbps.

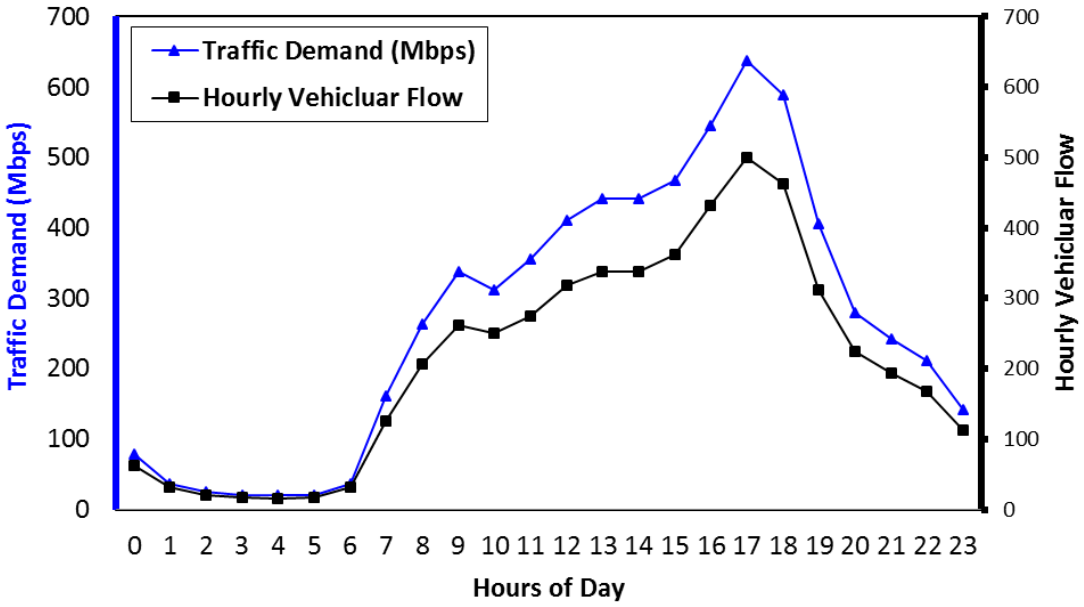


Figure 4.4: Hourly vehicular flow and traffic demand.

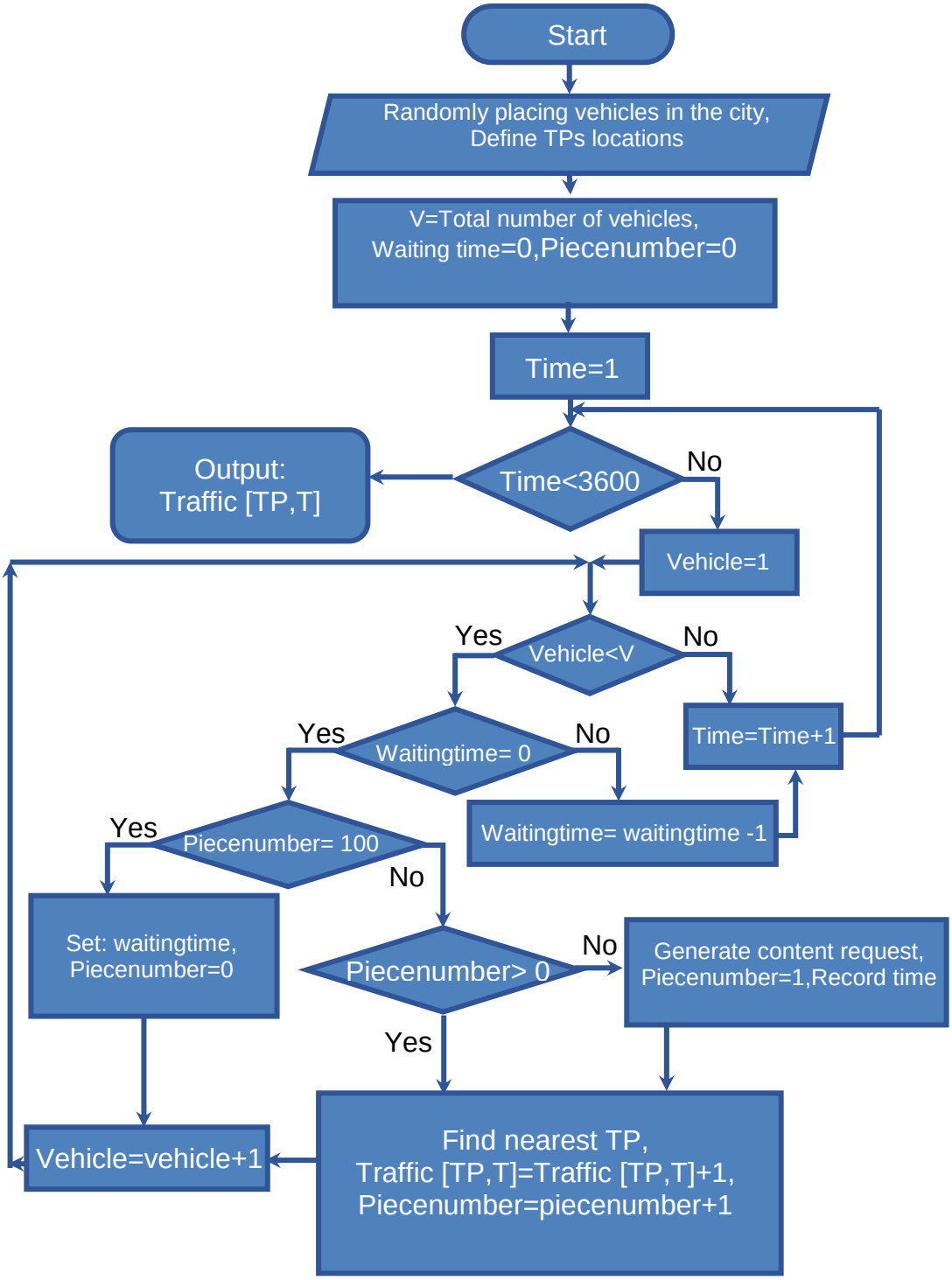


Figure 4.5: Traffic generation and traffic demand for each time point.

4.6 CPs number and location MILP optimisation

This study will focus on the possibility of having a Wi-Fi power-efficient Content Distribution Network in a city. For optimising the number of required CPs, we developed a mixed integer linear programming (MILP) model for a city vehicular network. A MILP model for minimising the total power consumption of a city's vehicular network by optimising the number of CPs while serving varying levels of traffic demand for different vehicular flows at each hour of the day has been developed. The sets, parameters and variables that are defined in Table 4.3 are used in the model.

Table 4.3 : List of Symbols.

Notations	**Description**
Sets:	
TP	Set of traffic points
CP	Set of caching point candidates
$N\{j\}$	Set of CPs at neighboring $TP\ j$
$N\{c\}$	Set of TPs at neighboring $CP\ c$
T	Set of time points within one hour i.e. 90
Parameters:	
$Pmax_{ct}$	Maximum power consumption of $CP\ c$ at time t i.e. $20\ W$ [98]

$Pidle_{ct}$	Operational power consumption of $CP\ c$ at time $t, Ceiling(P_{ct}\ /\ (1.3548) = 14\ W$ [99]
$Ptxmax_{ct}$	Maximum transmitter power consumption of CP c at time t , P_{ct} - $Pidle_{ct}$=6 W [99]
$Ptxmin_{ct}$	Minimum transmitter power consumption of $CP\ c$ at time t i.e. 1 W
Pw	Hourly available wind power
λ_{jt}	Traffic demand at $TP\ j$ at time t
$Ndmax_{ct}$	Maximum number of simultaneous downloads of $CP\ c$ at time t (i.e. 10)
$drmax_{ct}$	Maximum data rate at $CP\ c$ at time t (i.e. 30 Mbps)
$drmin_{ct}$	Minimum data rate at $CP\ c$ at time t (3 Mbps)
A	Large number for MILP constraint, set to 600 here
Variables:	
P_{ct}	Power consumption of a CP i.e. $14{\sim}20\ W$
Ptx_{ct}	Transmitting power consumption of CP c at time t

Pre_{ct}	The part of the available renewable used by CP *c* at time *t*
$Pnre_{ct}$	Non-renewable part of required power consumption of $CP\ c$ at time t
λ_{cjt}	Traffic between $CP\ c$ and $TP\ j$ at time t
dr_{ct}	Adaptive Data rate at $CP\ c$ at time t (3~30 Mbps)
α_c	Equals 1 if $CP\ c$ is on, equals 0 otherwise
β_{ct}	Equals 0 if Renewable energy is adequate for power consumption of CP, equals 1 otherwise
δ_{cjt}	Equals 1 if $CP\ c$ is transmitting content to Vehicle *j*, equals 0 otherwise
Nd_{ct}	Adaptive number of simultaneous downloads of $CP\ c$ at time *t* (i.e. $0\sim10$)
Indices:	
c	Index of caching point (CP)
j	Index of traffic points (TP)
t	Index of time point (T)

4.6.1 All Non-Renewable Sources with Non-Adaptive Capacity of CPs

The objective function of the optimisation is to **minimise** the power consumption of the active cache points which is given by:

$$\sum_{t\epsilon T}\sum_{c\epsilon CP}\alpha_c\,(Pidle_{ct}+Ptx_{ct}) \tag{4.10}$$

$$\forall\ t\epsilon T, \forall\ c\epsilon CP$$

The optimisation is subject to the following constraints:

$$dr_{ct}=(\,P_{ct}-\,Pidle_{ct})\frac{drmax_{ct}}{Pmax_{ct}-\,Pidle_{ct}} \tag{4.11}$$

$$\forall\ t\epsilon T, \forall\ c\epsilon CP$$

$$Nd_{ct}=\frac{dr_{ct}}{drmin_{ct}}=(\,P_{ct}-\,Pidle_{ct})\frac{drmax_{ct}}{drmin_{ct}}\,\frac{1}{Pmax_{ct}-Pidle_{ct}} \tag{4.12}$$

$$\forall\ t\epsilon T, \forall\ c\epsilon CP$$

$$\lambda_{jt}=\sum_{c\epsilon N[j]}\lambda_{cjt} \tag{4.13}$$

$$\forall\ t\epsilon T, \forall\ j\epsilon TP, \forall c\epsilon CP$$

$$\sum_{\mathrm{j}\epsilon\mathrm{N[c]}} \frac{\lambda_{\mathrm{cjt}}}{drmax_{ct}} \leq Ndmax_{ct} \quad (4.14)$$

$$\forall\ \mathrm{c}\epsilon\mathrm{CP}, \forall\ \mathrm{t}\epsilon\mathrm{T}$$

$$\frac{\lambda_{\mathrm{cjt}}}{drmax_{ct}} \leq\ \delta_{\mathrm{cjt}}\,\mathrm{A} \quad (4.15)$$

$$\forall\ \mathrm{c}\epsilon\mathrm{CP}, \forall\ \mathrm{j}\epsilon\mathrm{TP}\ , \forall\ \mathrm{t}\epsilon\mathrm{T}$$

$$\frac{\lambda_{\mathrm{cjt}}}{drmax_{ct}} \geq\ \delta_{\mathrm{cjt}} \quad (4.16)$$

$$\forall\ \mathrm{c}\epsilon\mathrm{CP}, \forall\ \mathrm{j}\epsilon\mathrm{TP}\ , \forall\ \mathrm{t}\epsilon\mathrm{T}$$

$$\sum_{\mathrm{t}\epsilon\mathrm{T}} \sum_{\mathrm{j}\epsilon\mathrm{N\{c\}}} \delta_{\mathrm{cjt}} \geq\ \alpha_{\mathrm{c}} \quad (4.17)$$

$$\forall\ c\epsilon CP, \forall\ \mathrm{t}\epsilon\mathrm{T}$$

$$\sum_{\mathrm{T}\epsilon\mathrm{T}} \sum_{\mathrm{j}\epsilon\mathrm{N\{c\}}} \delta_{\mathrm{cjt}}\ \leq\ \alpha_{\mathrm{c}}\,\mathrm{A} \quad (4.18)$$

$$\forall\ c\epsilon CP, \forall\ \mathrm{t}\epsilon\mathrm{T}$$

$Pidle_{ct}$ is the minimum operational power consumption of a CP while $CP\,c$ is switched ON that includes the power consumption of circuitry and receiver power consumption i.e. $14\,W$ [98]. We assume each CP has only ($1\,GB$) memory, which is of type small flash memory with near zero power consumption. Maximum power consumption of a CP including the memory is denoted as P_{maxct}. Ptx_{ct} is the transmission power consumption of a CP. We introduced simplistic (linear) relationship between transmission power of a CP and the data rate of CPs that the CP can offer (shown in figure 4.6) Data rate of a CP is given by equation (4.11).

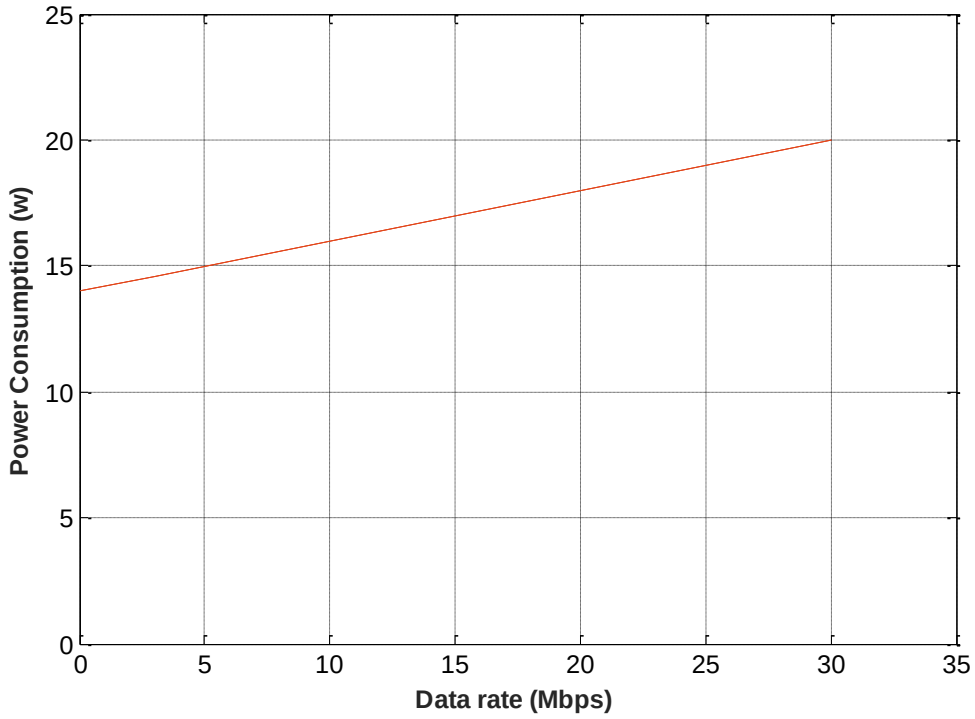


Figure 4.6 : Relation between Data rate & Power consumption.

The number of simultaneous connections of a CP from (4.11) can be defined as Equation (4.12).Equation (4.13) is the flow conservation constraint which ensures that at each time point total traffic is served. Equation (4.14) is the capacity constraint which ensures that at each time point the capacity of the CP is not violated, which is defined in terms of number of simultaneous downloads. CPs are non-adaptive capacities so data rate is always $drmax_{ct}$ and the number of simultaneous connections is maximum ($Ndmax_{ct}$). If the CP has already reached its full capacity, the model should install another cache to serve the remaining traffic of the TP.

Equation (4.15) ensures that if the traffic is non-zero between CP c and TP j i.e. $\lambda_{cjt} \neq 0$, then put $\delta_{cjt} = 1$. This is ensured by setting the constant A to a number, which depends upon the magnitude of traffic between a TP and a CP. In the present case, $A = 600$, which makes the right hand side comparable with the left hand side. Equation (4.16) ensures that if the amount of traffic between CP c and TP j is zero. i.e. $\lambda_{cjt} = 0$, then there is no connection hence, $\delta_{cjt} = 0$.

Equation (4.17) ensure that in the absence of traffic between CP c and all the traffic points in the neighbourhood (i.e. left hand side of the equation is equal to zero) α_c become, 0, (i.e. CP is switched off). Equation (4.18) switches on the CP, i.e. if there is a connection between CP c and TP j at time point t, the left hand side of the equation is greater than zero, then α_c can only be 1 to satisfy equation (4.18).

4.6.2 All Renewable Sources with Non-Adaptive Capacity of CPs

For this setup, we assume that the CDN does not have access to non-renewable power. Thus, the traffic demand can only be served, whenever the available wind power is sufficient. Otherwise, the CP would stay switched OFF. The objective is to minimise number of CPs to minimise energy consumption (equation 4.10). This MILP model is subject to the following constraints along with equations (4.13) to (4.16).

$$Pw = \sum_{t\epsilon T} \sum_{c\epsilon CP} Pre_{ct} \quad (4.19)$$

$\forall$ cϵCP, $\forall$ tϵT

$$Pre_{ct} \leq Pidle_{ct} + Ptxmax_{ct} \quad (4.20)$$

$\forall$ cϵCP, $\forall$ tϵT

$$Pnre_{ct} = Pidle_{ct} + Ptxmax_{ct} - Pre_{ct} \quad (4.21)$$

$\forall$ cϵCP, $\forall$ tϵT

$$Pnre_{ct} \geq \beta_{ct} \quad (4.22)$$

$\forall c\epsilon CP, \forall t\epsilon T$

$$Pnre_{ct} \leq \beta_{ct} A \quad (4.23)$$

$\forall c\epsilon CP, \forall t\epsilon T$

$$\sum_{t\epsilon T} \sum_{j\epsilon N[c]} \delta_{cjt}(1 - \beta_{ct}) \geq \alpha_c \quad (4.24)$$

$\forall c\epsilon CP, \forall$ jϵTP , $\forall t\epsilon T$

$$\sum_{t\epsilon T} \sum_{j\epsilon N[c]} \delta_{cjt}(1 - \beta_{ct}) \leq \alpha_c A \quad (4.25)$$

$\forall c\epsilon CP$, $\forall$ jϵTP , $\forall t\epsilon T$

$$Z_{cjt} = \delta_{cjt}(1 - \beta_{ct}) \quad (4.26)$$

$\forall c\epsilon CP$, $\forall$ jϵTP , $\forall t\epsilon T$

$$0 \leq Z_{cjt} \leq 1 \tag{4.27}$$

$$\forall c \epsilon CP \text{ , } \forall \text{ j} \epsilon \text{TP , } \forall \ t \epsilon T$$

$$Z_{cjt} \leq \delta_{\text{cjt}} \tag{4.28}$$

$$\forall c \epsilon CP \text{ , } \forall \text{ j} \epsilon \text{TP , } \forall \ t \epsilon T$$

$$Z_{cjt} \ \leq (1 - \ \beta_{ct}) \tag{4.29}$$

$$\forall c \epsilon CP \text{ , } \forall \text{ j} \epsilon \text{TP , } \forall \ t \epsilon T$$

$$Z_{cjt} \ \geq \delta_{\text{cjt}} + (1 - \ \beta_{ct}) - 1 \tag{4.30}$$

$$\forall c \epsilon CP \text{ , } \forall \text{ j} \epsilon \text{TP , } \forall \ t \epsilon T$$

$$\sum_{\text{t}\epsilon \text{T}} \sum_{j \epsilon N[c]} Z_{cjt} \ \geq \ \alpha_c \tag{4.31}$$

$$\forall c \epsilon CP, \forall \text{ j} \epsilon \text{TP , } \forall \ t \epsilon \ T$$

$$\sum_{\text{t}\epsilon \text{T}} \sum_{j \epsilon N[c]} Z_{cjt} \leq \ \alpha_c \ A \tag{4.32}$$

$$\forall c \epsilon CP, \forall \text{ j} \epsilon \text{TP , } \forall \ t \epsilon \ T$$

Equation (4.19) ensures the total renewable power consumed by CPs is equal to total available renewable power and Equations (4.20 - 4.29) are incorporated into the model to switch OFF the CPs whenever renewable energy is inadequate.

Renewable part of power consumption should not exceed the maximum required power of a CP (4.20) so $Pnre_{ct}$ from equation (4.20) and (4.21), we calculate non-renewable power required by a CP at time point t ($Pnre_{ct}$). Equations (4.22) and (4.23) set a binary variable (β_{ct}) to be able to ensure that the model does not consume any non-renewable power. If the renewable power consumption is lower than the maximum power required to switch ON a CP, $\beta_{ct} = 1$, else $\beta_{ct} = 0$. Therefore equations (4.24) and (4.25) ensure if the available renewable power is below the requirement, $CP\ c$ is switched off. Z_{cjt} to introduced to linearise equation (4.24) and (4.25). Equations (4.28) and (4.29) ensure if binary variable $\delta_{\mathrm{cjt}} = 0$ or $\beta_{ct} = 1$, Z_{cjt} is set to zero but do not ensure that Z_{cjt} is set to 1 if there is traffic ($\delta_{\mathrm{cjt}} = 1$) and renewable power is enough ($\beta_{ct} = 0$)so equation (4.30) is needed. Equations (4.24) and (4.25) can be re-written as Equation (4.31) and (4.32).

4.6.3 All Renewable Sources with Adaptive Capacity CPs

In the previous setup (4.6.2), there may have been instances where the traffic demand could not be served. However, by introducing a simplistic (linear) relationship between transmission power of a CP and the capacity of the CP can offer, the total traffic demand can be served by all renewable wind power. This, however, is achieved at higher number of CPs, where a lower capacity for CPs is adapted to compensate for lower wind power. Note that to install/switch

ON a CP, a minimum operational power (i.e.14 W [100]) is required, hence only the transmit power can be varied depending upon the available wind power. Since, in this setup, our aim is to serve traffic demand with only renewable sources, while installing/switching ON the minimal number of CPs, we maximise power transmission of each CP so that the maximum capacity of CPs (in terms of simultaneous downloads), dependent upon the available wind power, is achieved.

The objective function is to minimise the number of CPs, subject to meeting the traffic demand and minimising the power consumption. To support maximum download rate, the data rate of a CP is adapted by the transmission power available through wind energy. Therefore, the objective of this optimisation is given as:

$$\boldsymbol{Minimise}$$

$$\sum_{\mathrm{t}\epsilon\mathrm{T}} \sum_{c\epsilon CP} \alpha_c \left(Pidle_{ct} + Ptx_{ct}\right) \tag{4.33}$$

$$\forall c\epsilon CP, \forall\ t\epsilon T$$

Subject to:

$$Y_{ct} = \alpha_c\, Ptx_{ct} \tag{4.34}$$

$$\forall c\epsilon CP, \forall\ t\epsilon T$$

$$0 \leq Y_{ct} \leq Ptxmax_{ct} \tag{4.35}$$

$$\forall c \epsilon CP, \forall\ t \epsilon T$$

$$Y_{ct} \leq \alpha_c Ptxmax_{ct} \tag{4.36}$$

$$\forall c \epsilon CP, \forall\ t \epsilon T$$

$$Y_{ct} \leq Ptx_{ct} \tag{4.37}$$

$$\forall c \epsilon CP, \forall\ t \epsilon T$$

$$Y_{ct} \geq Ptx_{ct} - Ptxmax_{ct}\ (1 - \alpha_c) \tag{4.38}$$

$$\forall c \epsilon CP, \forall\ t \epsilon T$$

$$\sum_{\mathrm{t}\epsilon \mathrm{T}} \sum_{c \epsilon CP} \alpha_c\ Pidle_{ct} + Y_{ct} \tag{4.39}$$

$$\forall c \epsilon CP, \forall\ t \epsilon T$$

Equation (4.33) is non-linear because of multiplying two variables. Y_{ct} has been introduced to linearise this equation. Equations (4.35-4.38) are used to remove this non-linearity with linear equations.

Equation (4.36) ensures $Y_{ct}$is set to 0 when $\alpha_c = 0$. Equations (4.37) and (4.38) set Y_{ct} to Ptx_{ct} when $\alpha_c = 1$. Equation (4.33) is re-written as Equation (4.39).

In addition to equation (4.13) this optimisation has the constraints below. Equation (4.14) is re-written as Equation (4.40) to incorporate the effect of adaptive transmission power.

$$\sum_{j \epsilon N[c]} \frac{\lambda_{cjt}}{dr_{ct}} \leq Nd_{ct} \tag{4.40}$$

$$\forall\, c\epsilon CP, \forall\, t\epsilon T$$

$$Pre_{ct} = Pidle_{ct} + Ptxmax_{ct} \quad (4.41)$$

$$\forall c\epsilon CP, \forall\, t\epsilon T$$

Equation (4.40) ensures that at each time point the adaptive capacity of the CP is adequate. If the CP has already reached its full capacity, the model should install another cache to serve the remaining traffic of the TP.

Equation (4.41) ensures that the amount of non-renewable power consumption by CP *c* is equal to zero.

4.7 CDN Heuristic

A CDN heuristic to validate four different cases of the proposed models has been developed and presented as the flowcharts in figure (4.6) to (4.8). These cases are: 1) renewable energy with non-adaptive capacity CPs 2) renewable energy with adaptive capacity CPs where the traffic request (demand) at each CP is already calculated, the number and location of the CPs is optimised. 3) A heuristic which begins by checking the available renewable energy and available number of simultaneous connections (Downloads) and 4) A heuristic for Non-Energy adaptive case (RE + Non-ADP -CPs) where these heuristics make different decision according to study scenario is highlighted on flowcharts.

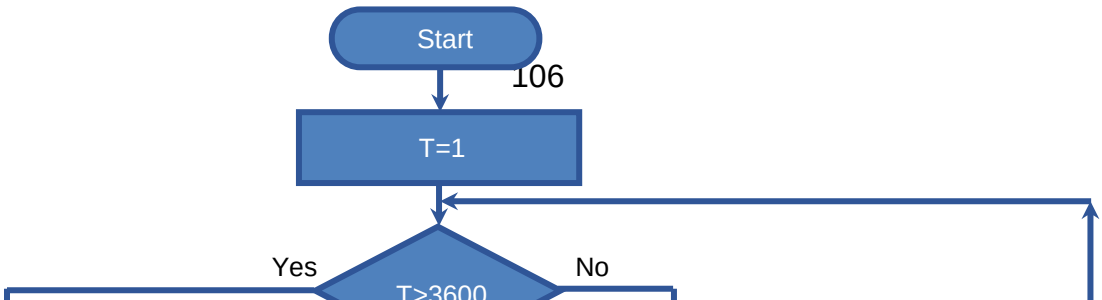
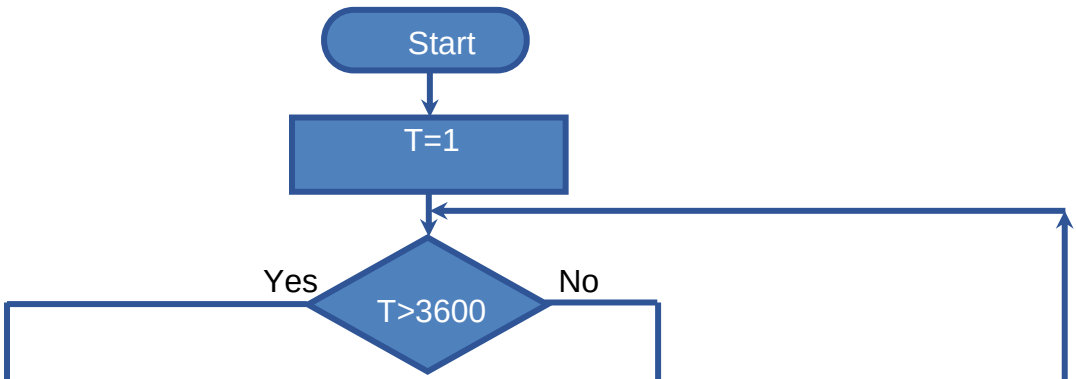
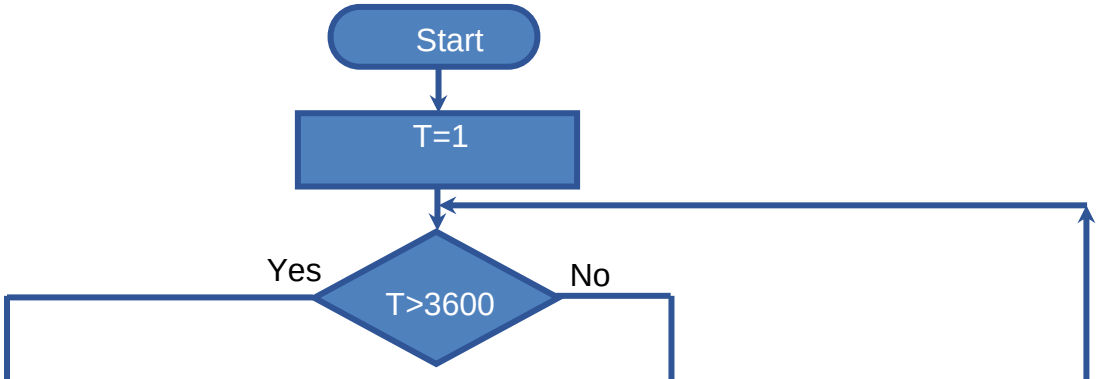


Figure 4.7: CDN Heuristic for Non Adaptive CPs with access to only non-renewable energy.




Start
T=1
Yes
T>3600
No

4.8 Performance Evaluation

From the results obtained by the MILP models and the heuristic for the CDN, we found the optimal number of CPs to minimise the total network power consumption according to the available renewable energy, and traffic. Figure 4.4 shows the total power consumption of the CPs for the proposed models. The result follows the same trend as the hourly vehicular flow Figure 4.2. The higher the vehicular traffic flow during peak hours the higher the total energy consumption of the CPs. In the case of renewable energy with non-adaptive CPs (RE+NonADP-CPs), all CPs get deactivated at the time periods when the wind energy available is less than the total energy required by the CPs. Further, it is observed that due to the lower capacity of the CPs in case of adaptive CPs, there would be an increase in the number of active CPs for serving traffic demand with the maximum download rate. Hence, the power consumption is higher compared to the other models.

Figure 4.10 shows the available capacity of CPs in different cases. According to the availability of wind energy, the CPs adjust the serving capacity. In case of NRE, the CPs utilise capacity due to the availability of sufficient wind energy. The figure illustrates that for the RE+NonADP-CPs, capacity is equal to zero due to inadequate renewable energy in some hours.

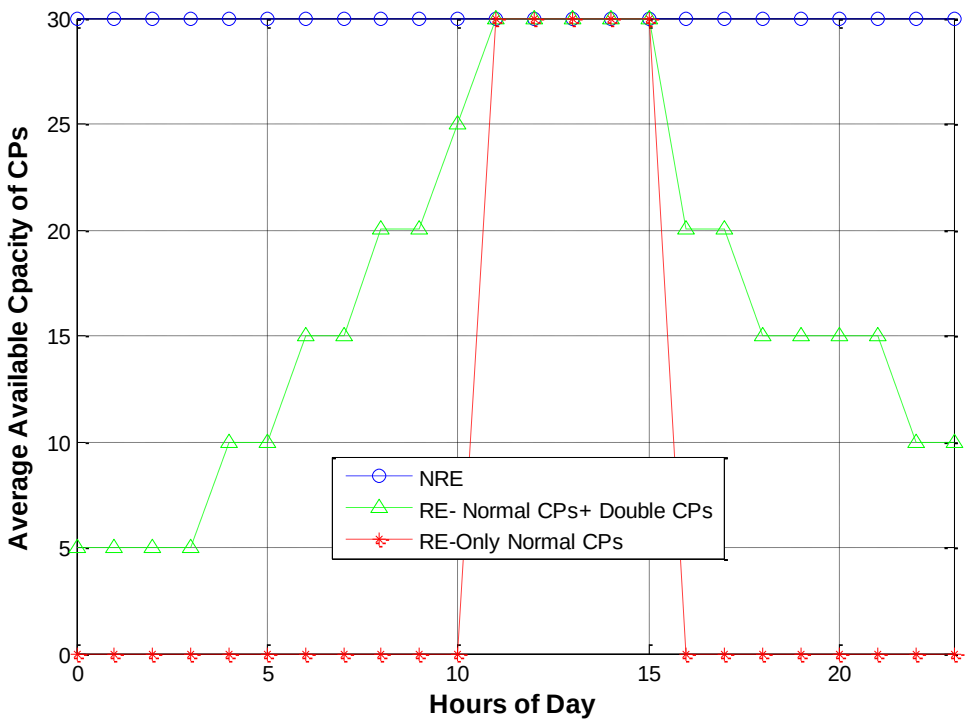


Figure 4.10 : Available capacity of CPs for varying hours of the day.

Figure 4.11 shows the total number of active CPs in the network obtained by respective models every hour of the day. The number of CPs needed to serve the content demand varies during the day. The number of active CPs is directly proportional to the traffic demand which is trivial. The results show that for the non-renewable energy model without energy saving (NRE), the number of CPs activated remains fixed during the lower traffic periods (00:00 – 06:00) as at least 12 CPs serve the demand. Whilst the traffic demand (after 06:00) increases, the number of CPs increases as well. For all renewable sources with only normal CPs model, while CPs are experiencing a deficit of available

renewable energy, the average number of active CPs is equal to zero. Figure 4.12 shows that the model with all renewable sources and adaptive capacity CPs requires switching ON of additional CPs during a shortage of renewable energy, even if the CPs serve at maximum download rate.

Figure 4.8 illustrates the total energy saving of each model in comparison with the initial model (*NRE-all CPs ON*). The energy saving is calculated as:

$$Saving = \frac{Pmax(\text{number of active CPs}) - \text{TotalAdaptedpowerconsumption}}{Pmax(\text{number of active CPs})} 100\%$$

It reveals that the energy saving by using renewable energy where by adding more active CPs, 100% saving is achieved. In the case of *NRE* with energy saving MILP, a sharp peak in energy saving at 4:00 am is observed, which is due to the lowest number of active CPs required at that time, reflecting lowest vehicular traffic.

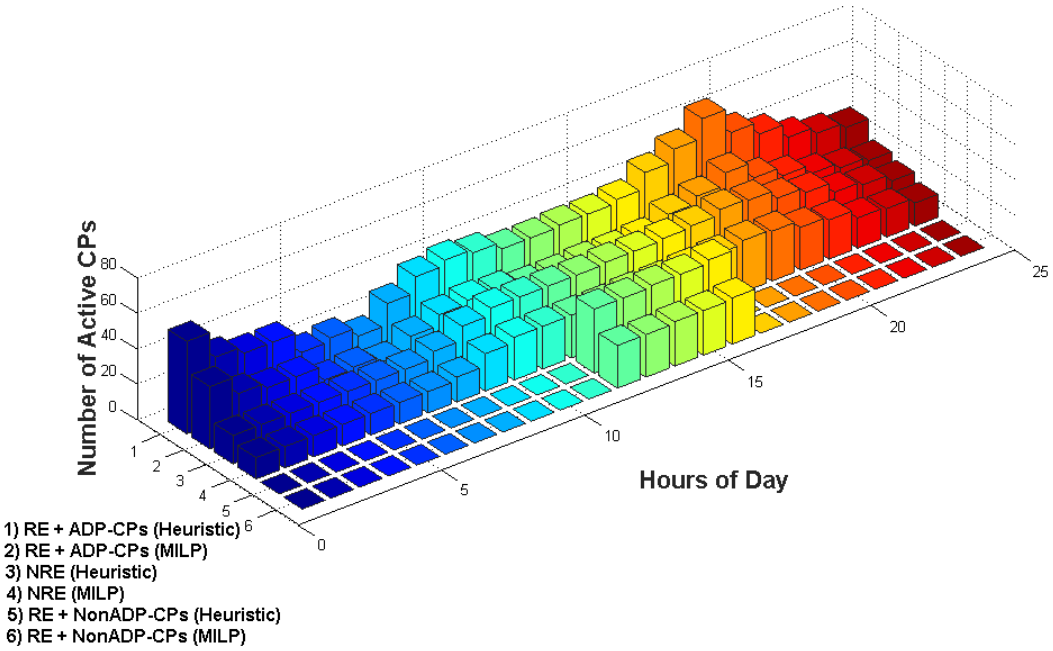


Figure 4.11: The number of active caching points for varying hours of the day.

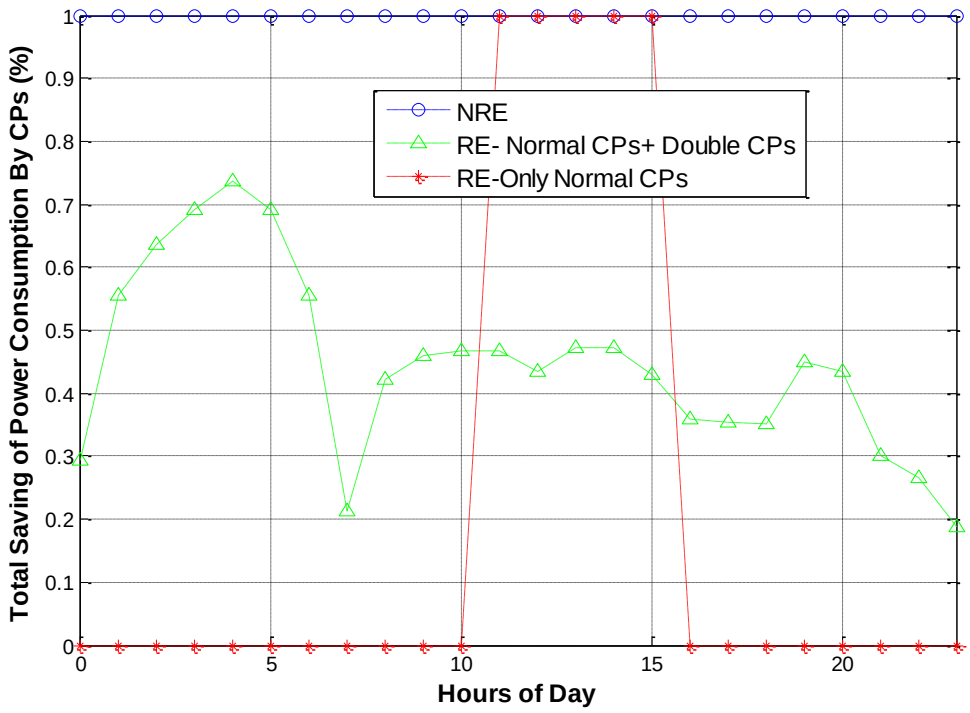


Figure 4.12 : Total power savings for varying hours of the day.

Figures 4.13 shows average piece delay in *NRE*, *RE-Only Normal CPs*, *RE-Normal CPs + Double CPs*. In the case of *RE- Normal CPs + Double CPs*, the average piece delay is higher. However, all energy consumption is renewable energy as the model tries to serve the remaining traffic demand due to the capacity (adaptive) bounds on active CPs. In the case of *RE + Non-ADP-CPs*, the system can only serve 4 hours; hence the respective delay is calculated. The figure shows the inability of *RE + Non-ADP -CPs* in powering up the CPs required to serve all traffic demand.

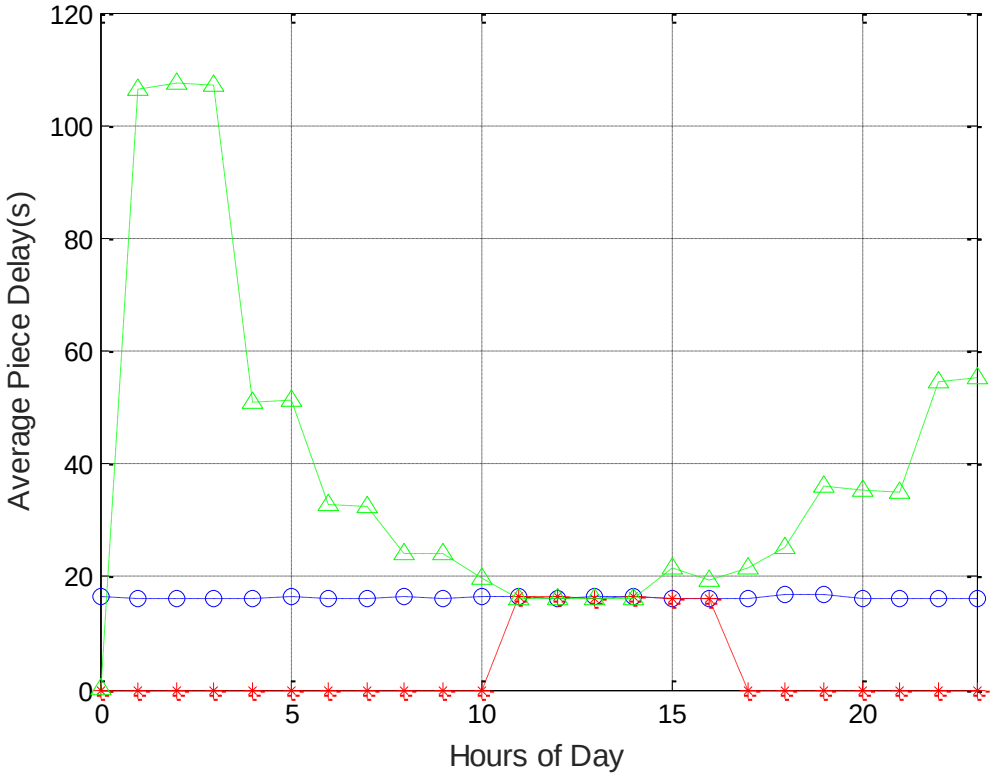


Figure 4.13 : Average Piece Delay for varying hours of the day.

Similar behaviour of average piece delay can be seen in average content delay (Figure 4.14), however with a lower magnitude. This is because inter-piece time (IPT) is the same in all cases, which allows the system to have more simulataneous downloads.

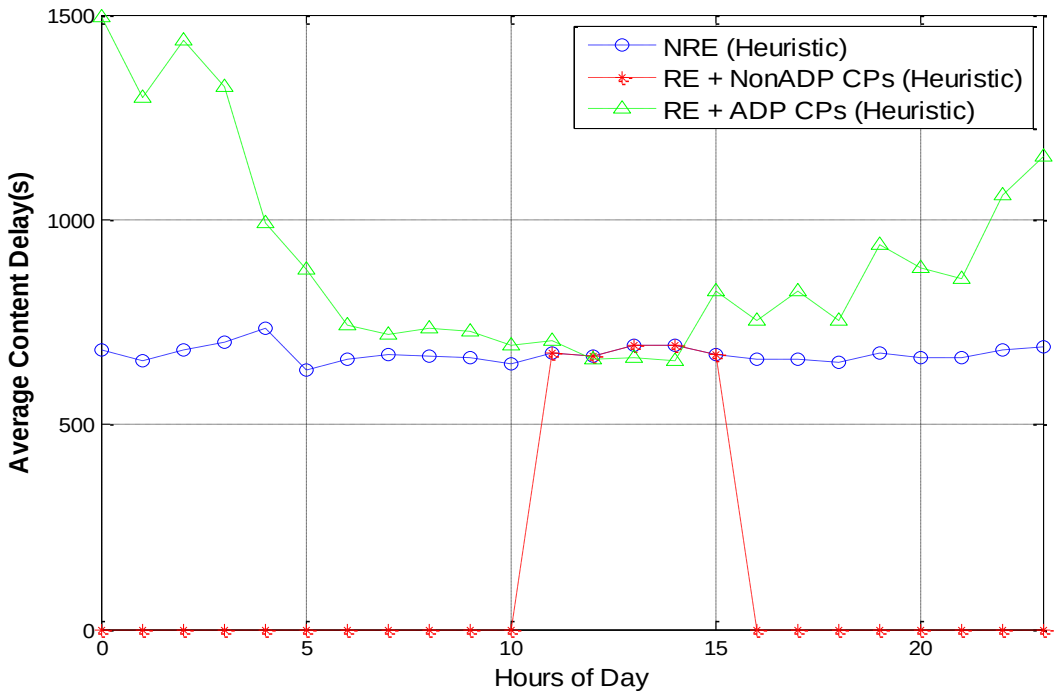


Figure 4.14: Average Content Delay for varying hours of the day.

4.8 Conclusions

In this chapter, energy adaptive CPs for vehicular CDN are proposed. The average piece delay at the proposed CPs was dependent on the available renewable energy. For non-adaptive CPs, the whole traffic demands can only be served at the expense of higher number of active CPs. This means that if we want to keep the same number of CPs, the whole traffic demand will not be served at hours. This increases the overall average piece delay at earlier hours of the day when the available wind energy is insufficient. In contrast, the non-renewable energy consumption increases at lower renewable energy due to the number of active CPs. The advantage of the proposed adaptation capacity technique at a CP operating only with renewable energy is evident in terms of acceptable average piece delay and highest non-renewable energy savings compared to that of a traditional CP. The performance results revealed that the proposed CP saved 100% non-renewable grid energy during the whole day while fulfilling content demand in a city vehicular environment.

In the next chapter, we propose a suitable solution for efficient smart city vehicular environment.

Chapter 5: Location Optimisation of Grid connected Nano-Servers in a City environment using alternative optimisation methods

5.1 Introduction

Due to the unprecedented growth in the infotainment segment, the industry is moving towards the service based abstraction for reducing capital expenditure and maximising profit share. Parallel growth in cities Gross Domestic Product (GDP) drives researchers towards the paradigms of smart city [100] to make better use of public resources, increase the quality of services offered to the citizens and travellers while reducing the operational cost of public administration [103]. Smart cities utilise different catapults. To name a few, these are smart grids, intelligent energy management [104], health and safety, smart traffic signalling and management, mobile infotainment. Such conglomeration would help creating better cities for human beings where objects around us know what we like, what we want, what we need and its

availability [105, 106]. In the information and communication domain it gives rise to the concept of Internet of Things (IoT) [103], which connects millions of devices together. Evidently, it requires modernisation of information and communication technology solutions in the mobile and vehicular domains. In a smart city, users and smart things gather information through communication and processing devices such as mobiles, iPad, WiFis (public hotspots) which are connected to the cloud infrastructure. Cloud infrastructure retains data longer but the upload is time consuming. The infotainment data is dynamic in nature, thus it is largely affected by mobility, number of users and real time applications. To deal with such dynamic nature, the radio range needs to have high bandwidth (and hence shorter distance). To solve such issue, intermediary devices (between cloud and end users) are needed. These devices in the current context are called Nano Servers (NSs) [49]. In smart-cities, vehicular users play a crucial role in road safety and pollution, thus it is of paramount importance to study the performance of Nano Servers serving vehicular traffic. Currently, the information and communications technology (ICT) sector contributes to 2%-2.5% of the globally emitted carbon, where this figure is expected to increase considerably in near future [84]. Therefore, energy efficiency in the nano Servers is of concern.

In this chapter, a MILP model is developed which minimises the total power consumption of the fog computing nano servers in smart city vehicular environment by optimising the number and locations of ordinary (non- adaptive)

NSs thereby reducing the overall network energy consumption. Secondly, a heuristic is developed to verify the results of the MILP model that mimics the behaviour of the proposed MILP model. Finally, the performance of the scenario is evaluated in terms of the number of installed and active nano servers and total power consumption. Then for validation of our proposed model we developed several numerical algorithms.

5.2 Smart City Vehicular Scenario

The vehicular scenario considered in this chapter is a smart city, where vehicle movement follows a Manhattan Mobility Model [107]. The city area considered is 3x3 km, where there are 24 bi-directional roads and 16 junctions as shown in Figure 5.1. There are a total of 398 possible locations (at roadsides and road junctions) i.e. candidate sites (CS) whereas the MILP model is unable to handle the movement of vehicles, we assume that a set of Traffic Points (TPs) represents traffic centroids which account for the content/data traffic requested by nearby vehicles. Hence, a set of fixed traffic-generating points in the city space is seen by the MILP model. Vehicular mobility (and vehicular data generation) causes the level of traffic (in Mb/s) at each of these traffic centroids to vary. It means that if a vehicle moves from one TP to another TP, the numbers of the vehicles of the corresponding TPs change. If a large number of such traffic points / TPs is adopted, the impact of discretisation is reduced. In the simulations we include a scenario where such discretisation is completely removed; vehicles roam freely

and generate traffic continuously. The results show good agreement with our discretisation approach, where the latter enables MILP modelling. Each bi-directional road has one TP, amounting to a total of 112 TPs in our city vehicular network. Similarly a set of 398 candidate sites (CSs) has been assumed to select a number of active CPs at any point in time. Since the city roads are narrow, the width of the roads is not considered. We increased the number of traffic points (TP) and candidate sites (CS) compared with Chapter 4 to give opportunity to each TP to have access to a larger number of CSs in the range and also to have CSs which are in the range of more than one TP. We consider short range communication for IoT content delivery for which the servers can potentially be placed 100 m apart. A section of the city is shown in figure 5.3. A vehicle can connect with any of the NSs in range (NS1 to NS6) [refer to figure 5.3]. Thus, the maximum range for connectivity is 200 m. We consider wifi for the communication between the vehicles and NSs and the vehicles are perfectly power controlled. This enables overlapping coverage of the NSs. In a typical UK city (e.g. Leeds) the speed limit is 30mph (+3 mph tolerance) in the majority of locations i.e. 15m/s. Thus even if a vehicle travels at this limit, the time to go out of range of the neighbouring NSs (dwell time) is 13 s.

Smart things in the city (IoT objects) such as commercial places, healthcare and educational buildings, and petrol stations are connected to NSs. The NSs are connected to the external network via cloud as shown in Figure 5.2. The IoT information piece contains data about the smart city such as environmental

monitoring data, flyers, bed and breakfast information, car park, service station information, medical and healthcare systems information, short movie trials, advertisements etc. A nano server may process such pieces to extract knowledge from the data and then shares such knowledge with the vehicular. A vehicle requests an IoT content from any of the neighbouring NSs in a piecewise fashion to avoid the complications associated with handoff management, such that a piece should be downloaded while the vehicle is in IoT domain [within the dashed box as shown in Figure 5.3]. Thus, we consider the IoT piece size to be 2 MB . IoT piece is a package of several objects with different sizes is considered for example a package of 1) five 3-page word processor document (18 kB each) like city information booklet/ Parking guidance 2) Two 1280x960 JPEG photos like restaurant menu 3) One minute audio as MP3 like new movies or theatres' advertisement). The NS can accept a maximum of 9 connections, which implies that each IoT piece would be downloaded with 3 Mbps data rate (at least) out of 27 Mbps total bandwidth. Thus, the download time is 5s, which is less than the time required to traverse (13 s) the dashe box of Figure 5.3.

Figure 5.1: Smart City Vehicular Environment.

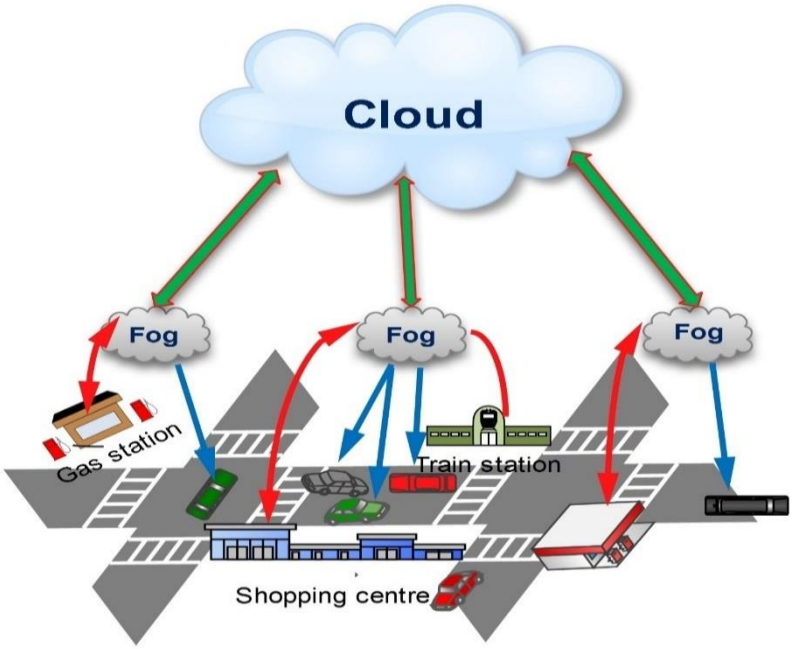


Figure 5.2 : Fog computing architecture diagram for IoT delivery in Smart City scenario.

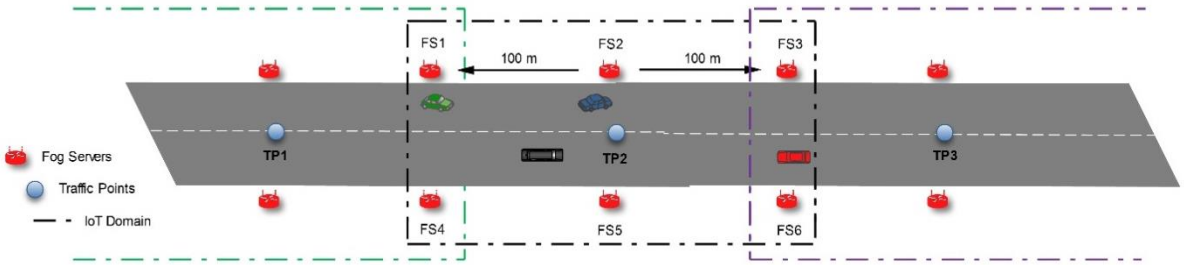


Figure 5.3: IoT domain and coverage range of each nano server.

5.2.1. Power consumption profile of a Nano Server (NS)

The power consumption model of the Nano server is described as follows. In our case, Fog computing servers are nano-servers, each having central processing unit (CPU), Memory and Disk. The minimum operational power consumption of a Nano Server comprises of full CPU power plus half of power usage of RAM and hard disc. The other half of power usage is for transmission. From [108], we obtained the percentage energy usage of CPU, memory, and hard disk, which are 58%, 28% and 14%, respectively. Since the total power consumption is 30 W [109], the minimum operational power consumption is 23 W. Therefore, the transmission power consumption is (30 - 23 = 7 W). The computing services consume CPU resource (cycles). Further, 7 W is the maximum transmission power consumption (consisting of networking and some CPU resource). The data rate of the rate adaptive NS linearly varies according to the available transmission power [110]. Without rate adaptation, the NS consumes 30 W and operates at a maximum data rate of 27 Mbps [111]. The NSs are grid connected and therefore

have access to both NRE and RE. A detailed discussion can be found in Chapter 4.

5.3 MILP Model for Transient Traffic based Nano Server Location Optimisation

As mentioned in Section 5.2, there are a total of 398 possible locations i.e. candidate sites (CS) where NSs can be installed. We developed a mixed integer linear programming (MILP) model which chooses hourly sets of optimised locations CS_t^{opt} (i.e.$CS_t^{opt} \subseteq CS$) from all possible locations (CS). The MILP model considers vehicle mobility and data traffic transiency through TPs to yield time dependent optimised locations of the NSs (transient optimisation). The union over all optimised locations would give rise to the set of active CS, where NSs need to be installed. This also minimises the overall power consumption of the NSs. Note that the NSs here are non-rate adaptive, which means that these operate at full power regardless of the traffic. We obtained this optimisation results from MILP model in Section 4.6.1 in Chapter 4.

Although, the professional (business) version of the AMPL can accommodate 3 million variables and constraints, it is not surprising for CPLEX to run out of memory in case of transient level optimisation because of the sizes of the branches & cut trees formulated in each Linear Programming sub-problem.

CPLEX saves a basis for every unexplored node. Unfortunately, once a solution process has failed because of insufficient memory, the amount of additional memory required for final solution cannot be proactively determined [112]. The MILP Model 1 yield 44576 variables per second. Thus, for a time period of 600 s, the transient optimisation problem yields 26,745600 variables. Consequently, CPLEX runs out of memory during the solution process. To alleviate this limitation, we run the model for every 60 time points separately (which keeps the number of variables within limit) and find out 10 different set of optimised locations (and number) of NSs. A union over all the sets gives us a new set of CS, which is much smaller than the initial set of CS. This ensured that the number of variables remain within the acceptable limit of CPLEX for the entire duration of 600 s. We re-run the MILP using the new set of CS to obtained optimised number and locations of the NSs.

5.4 Other Optimisation Methods

The MILP model in Chapter 4 for CP location optimisation was verified only through a heuristic approach. In this chapter, we introduce 3 other optimisation methods (Bisection, Greedy and Genetic) and use them to verify the MILP model.

5.4.1 Custom Heuristic

We initially developed a custom heuristic method for our MILP model with new parameters. The flowchart shown in Figure 5.4 describes our heuristic. We defined the traffic for each traffic point and at each time point to be the same as the MILP input to allow comparison to MILP. The main components of this heuristic are:

1. A central Controller is introduced
2. For traffic point j before switching on a new NS, the controller checks traffic at the rest of the TPs. If any other TP, neighbouring TP[j], has traffic at that time, the NS in the range of both TPs is turned on.
3. If more than one NS is needed, Priority is given to the NS which is in the range of most TPs. CP3 are CPs in range of three TPs, CP2 are CPs in range of two TPs and CP1 are only in range of one TP.

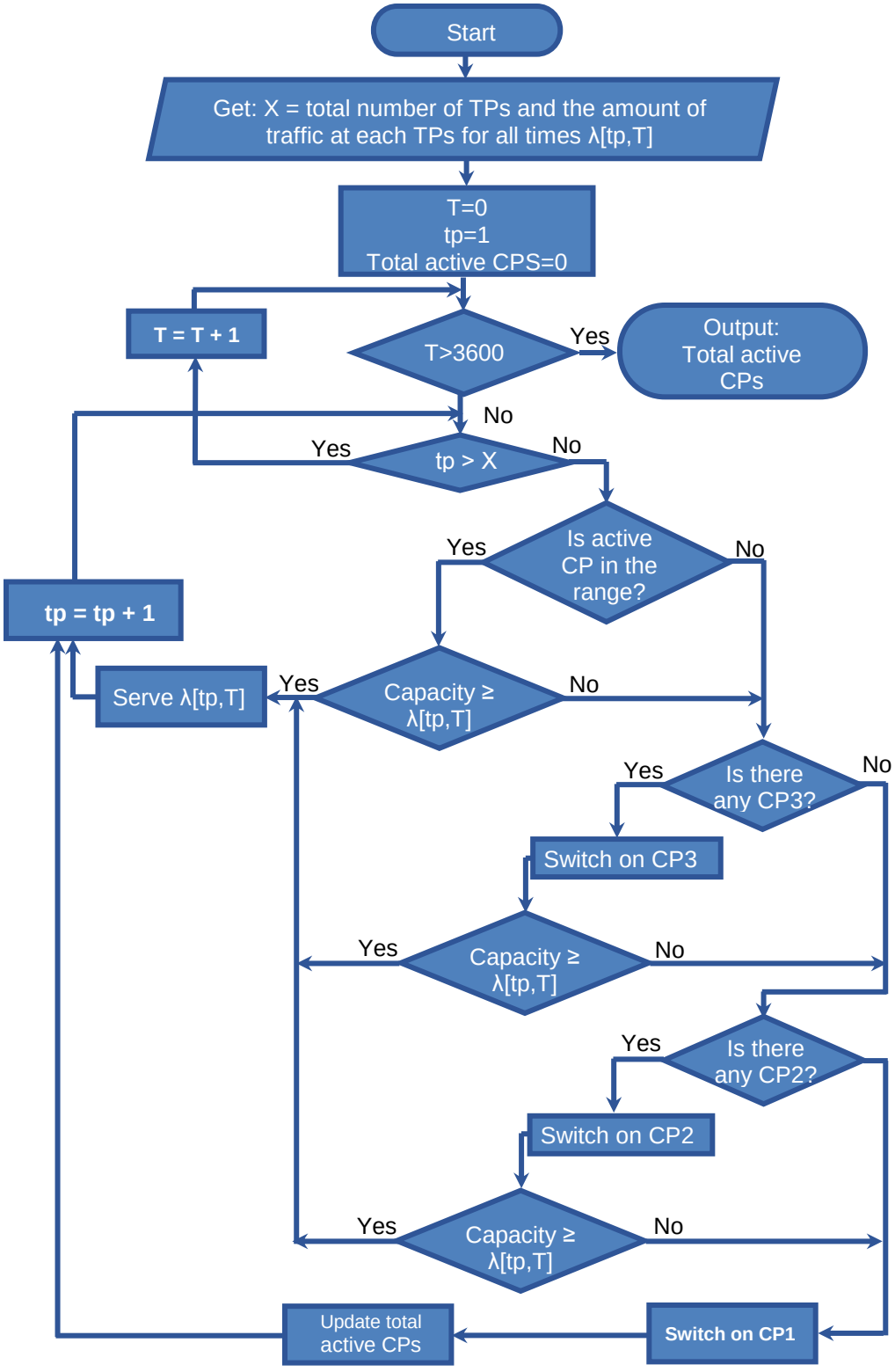


Figure 5.4: Location Optimisation with proposed custom heuristic.

5.4.2 Bisection Algorithm

The bisection method in mathematics is a root-finding method that repeatedly bisects an interval and then selects a subinterval in which a root must lie for further processing. It is a very simple and robust method, but it is also relatively slow. Because of this, it is often used to obtain a rough approximation to a solution which is then used as a starting point for more rapidly converging methods [113]. The method is also commonly known in the literature as the interval halving method [114], the binary search method, [115] or the dichotomy method [116]. Bisection defines a process (bisecting) and use this process to reach the answer, it may be slow but will get the answer. The bisection location optimisation algorithm developed is shown in Figure 5.5. First, set up an array which has as many as elements as there are NSs.

Randomly generate 2 values where each of them is an array that includes some of the NSs cells. The condition required is that a is able to serve the maximum half of the traffic and b needs to serve all the traffic. Bisect interval, i.e find an array, c, with size exactly in the middle of the size of a and b. c is populated with elements from all the elements of $a \cup b$. Next, the amount of traffic that c can serve ($f(c)$) is determined. If $f(c)$ is higher than $f(a)$ replace a by array c otherwise b is replaced by c and the process continues bisecting till the difference between a and, b is less than 0.05 and a is able to serve the total amount of traffic, a is the optimum number and locations of NSs, otherwise b.

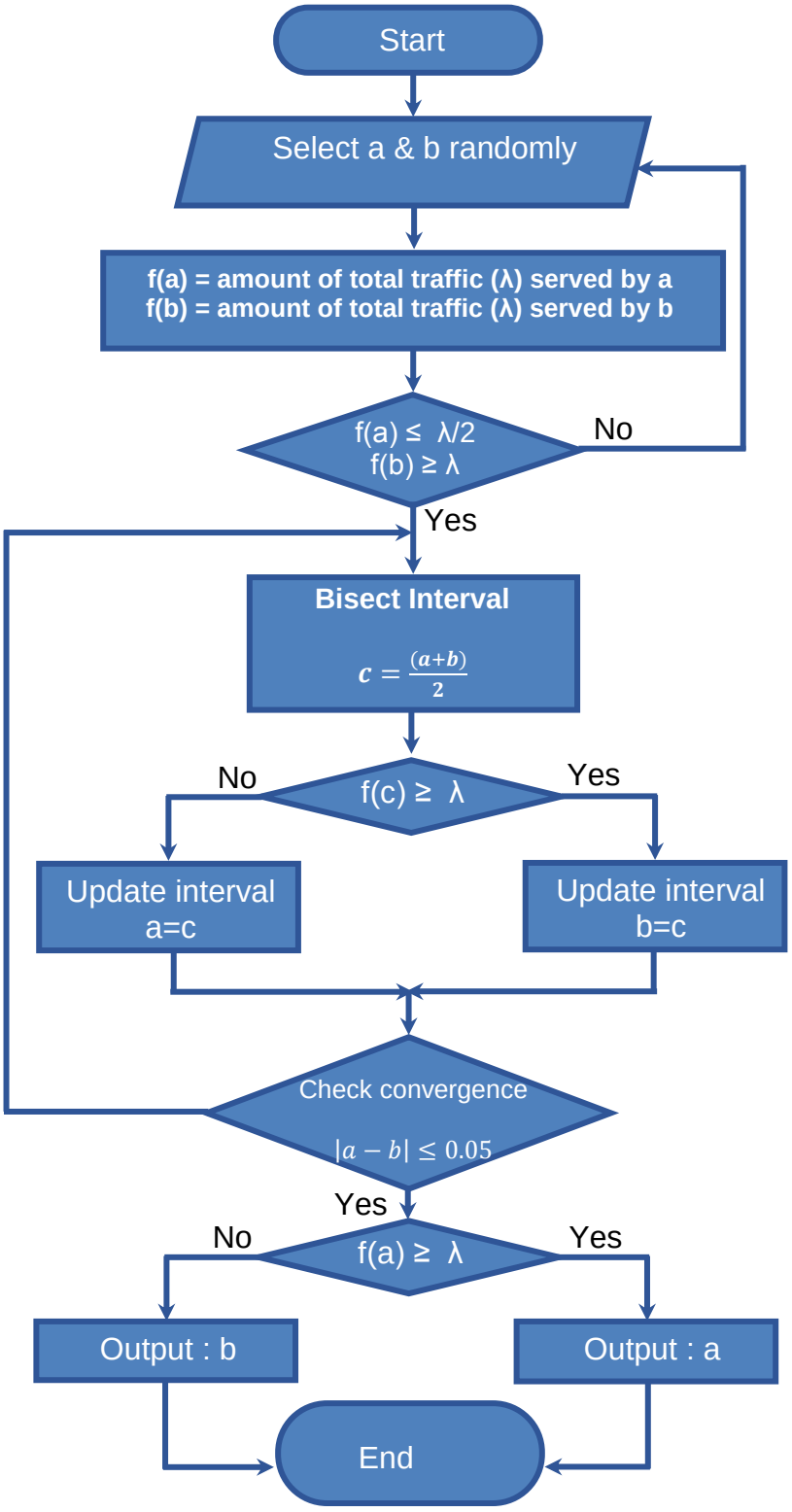


Figure 5.5: Location Optimisation with bisection algorithm method

5.4.3 Greedy Algorithm

A greedy algorithm is a mathematical process that looks for simple, easy-to-implement solutions to complex, multi-step problems by deciding which next step would provide the most obvious benefit. Such algorithms are called greedy because, while the optimal solution to each smaller instance would provide an immediate output, the algorithm doesn't consider the larger problem as a whole. Once a decision has been made, it is never reconsidered. Greedy algorithms work by recursively constructing a set of objects from the smallest possible constituent parts. Recursion is an approach to problem solving in which the solution to a particular problem depends on solutions to smaller instances of the same problem. The advantage of using a greedy algorithm is that solutions to smaller instances of the problem can be straightforward and easy to understand. The disadvantage is that it is entirely possible that the most optimal short-term solutions may lead to the worst possible long-term outcome. Our developed greedy algorithm is described in Figure 5.6. There is an array (NSs) with 398 elements that represent all the NSs. Then n sub-arrays are generated where the size of each array is randomly selected between 1 and 398 (i.e. $n = 100$). The requirement is that at least half of the sub-arrays ($n/2 = 50$) should serve all the traffic (λ). If the total arrays that can serve the whole traffic is less than half, then for each array add one more element randomly from NSs and check it again. This process continues till this array satisfies the full traffic. After

we have 50 arrays that all satisfy the total traffic, stop and select the array that results in the minimum power consumption.

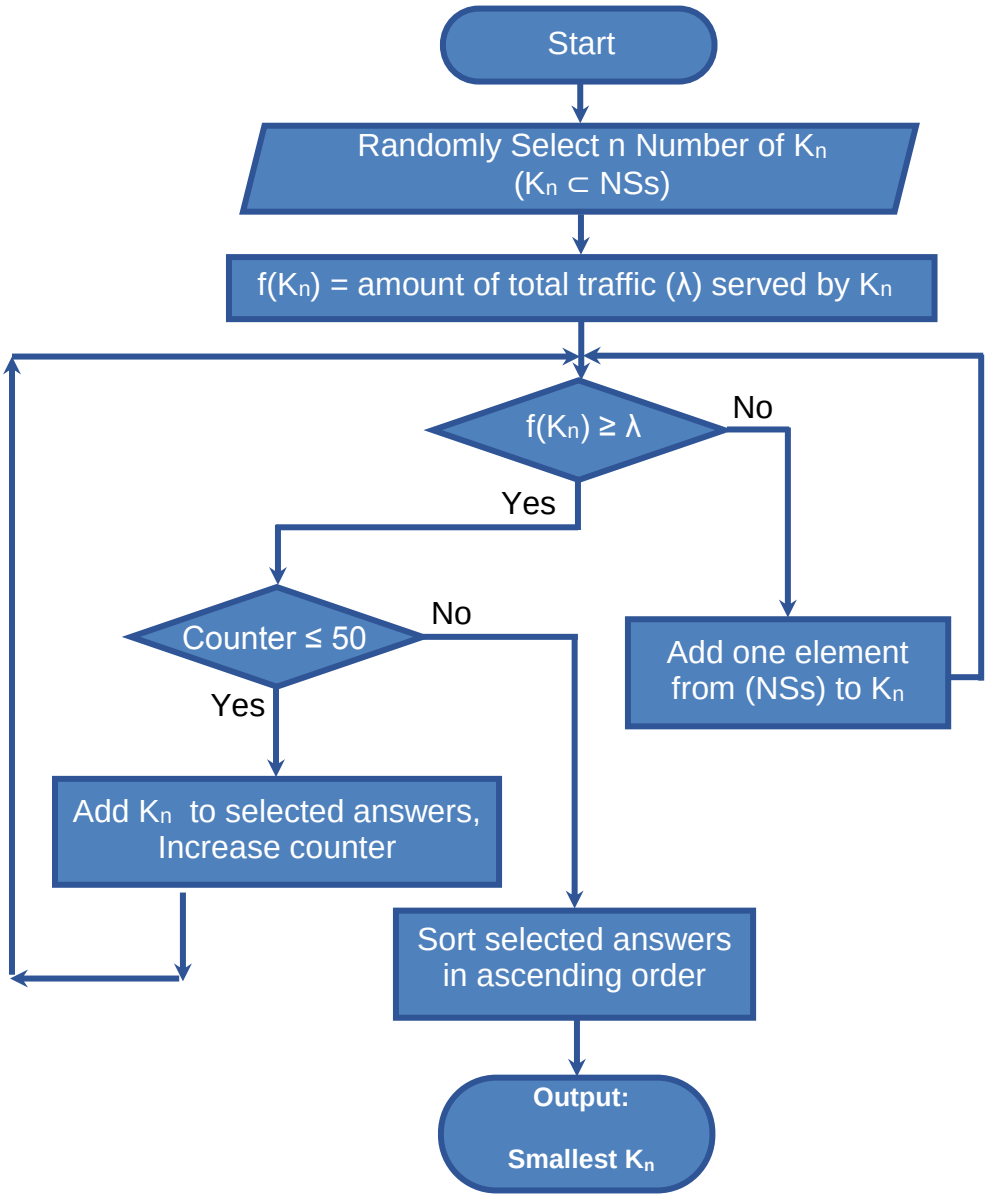


Figure 5.6: Location Optimisation with greedy algorithm method.

5.4.4 Genetic Algorithms

Genetic Algorithms (GAs) are one of the most popular heuristic algorithms to solve optimisation problems. Genetic Algorithms are inspired by natural evolution to find the optimum solution to a range of problems. Evolution is the process that allows different species to improve their chances of survival in the wild environment. It is through evolution that these species adapt to changes in their environment. As species evolve they can become more complex and significantly different from their previous generations. Evolution works by incorporating changes, usually small, that make the species stronger and fitter and removes attributes that do not improve fitness. This process works because individuals with advantageous attributes survive longer in their environment to reproduce, whereas individuals without the attributes or with worse attributes die earlier or are unable to attract mates [134].

The key elements of a genetic algorithm are:

- **Genome:** An encoded set of parameters
- **Individual:** Each member of the population competing for survival. Each individual consists of at least a genome.
- **Genetic Operator:** A function that takes a number of genomes and produces a single genome.
- **Fitness Function:** A function that determines the fitness level of each individual and subsequently their suitability as the solution to the problem.

The genetic operators used in our experiments are:

- Selection
- Crossover
- Mutation

From three main classifications of crossover operators (Single Point, Multi Point and Uniform). We used uniform crossover operator in our genetic algorithm for location optimisation. Parents with the same size (number of NSs) are selected. Then, each gene of the new genome is picked from one of the parents at random.

¦ = Break point

Parent 1: 0¦00¦00¦000¦0¦0

Parent 2: 1¦11¦11¦111¦1¦1

Child: 0110011101

The second operator used by our GA is the mutation operator. This operator takes a genome and then mutates its genes randomly. For a binary string, each bit has a small chance of being inverted. The probability of each bit being inverted is called the mutation rate and is normally set at a level that can create diversity but not too high to distract the algorithm from the solution. An example of mutation at bits 5 and 8 is as follows:

0123456712 ⟶ 0123956812

A generic algorithm for the proposed location optimisation model has been developed and is presented in flowchart 5.4.

The proposed Genetic Algorithm begins by generating the initial population randomly, which in present case is 400 set of optimised locations of the NSs. Each pair of chromosomes is of equal size but the sizes are random between 1 and 398. The next step is the selection operation. The top half chromosomes from the population based on the amount of traffic which they can serve are selected to pass on their genes to the next generation. Crossover needs to be done on these selected chromosomes and one offspring is kept to add to the top selected population. Mutation needs to be done at the same time applied to 10% of the population. Mutation in our genetic algorithm is random in 1 gene and then mutated chromosomes replace the existing chromosomes so the population size remains the same. Every iteration the chromosomes are sorted by the amount of traffic they can serve until the stop criterion is satisfied, the GA checks the size of each chromosome and output the chromosome with the smallest cost to minimise power consumption.

The iteration results are displayed in Figure 5.5, which shows improvement of energy saving (Fitness) for hour 12:00. The effect of generation size (iteration) on the fitness values suggests that running the algorithm for longer than 650 gives only a trivial amount of improvement and after 700 iterations remains at a saturation level. The genetic validation results are very robust and validate our MILP model well (4% deviation). The developed algorithms are compared in Table 5.2. As can be seen in this table, the genetic algorithm has the closed results to our MILP results.

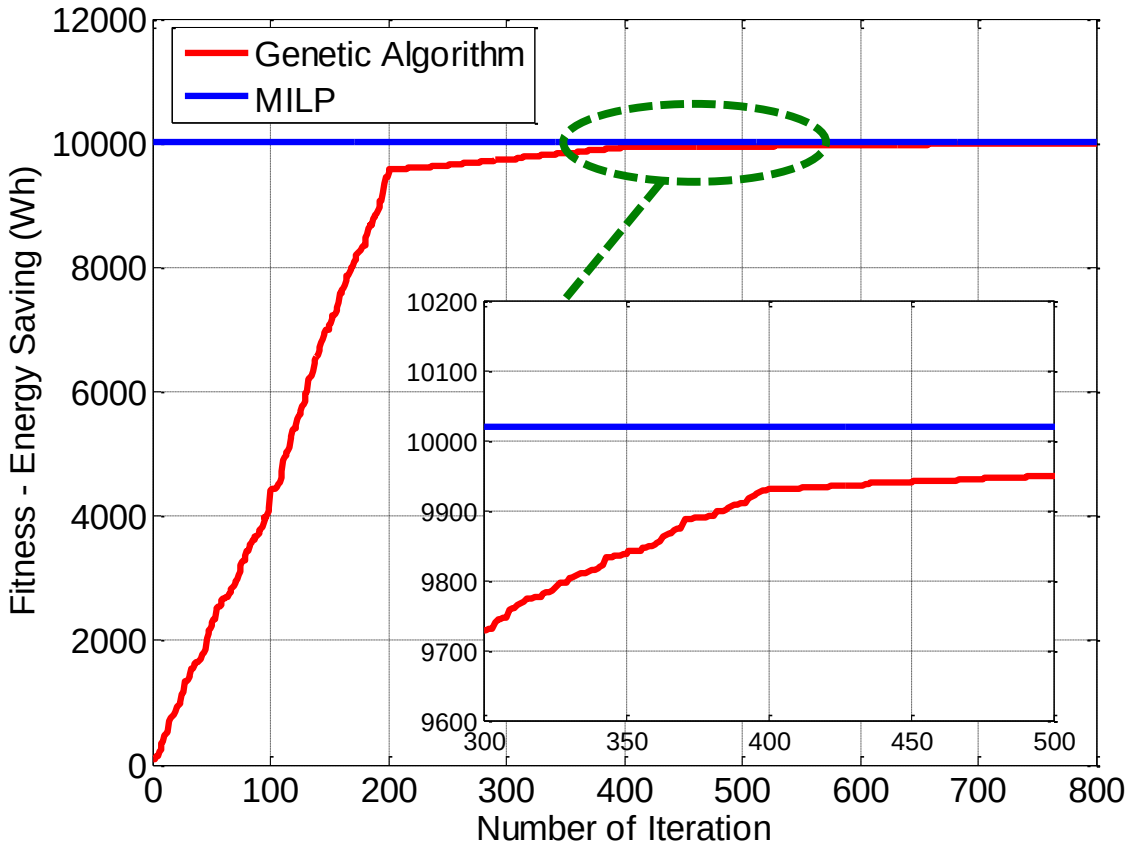


Figure 5.7: Fitness value for different number of iterations on proposed genetic algorithm.

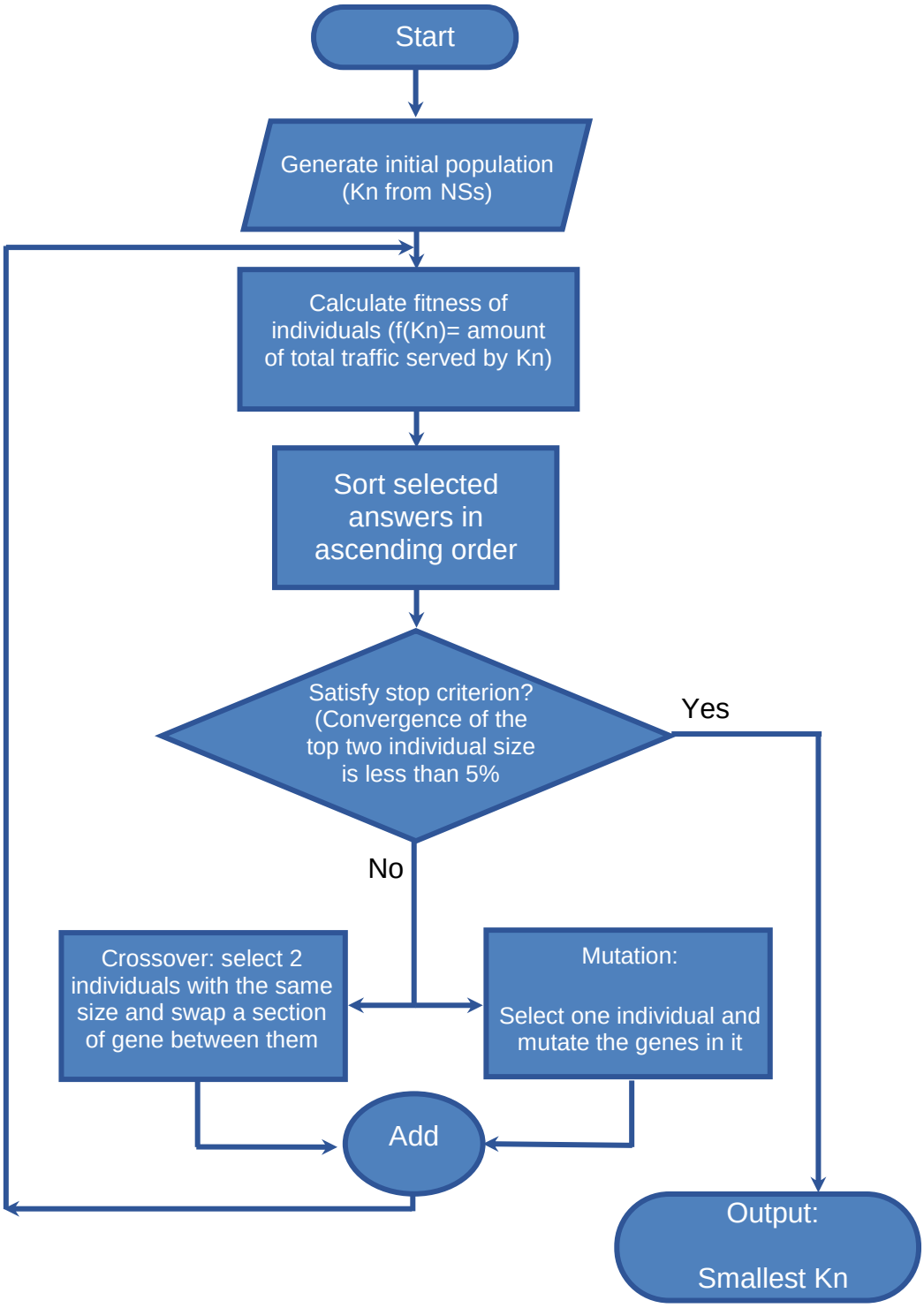


Figure 5.8: Location Optimisation with genetic algorithm method.

Table 5.1: Location Optimisation Algorithms.

Algorithms	Complexity	Deviation % from MILP
Custom Heuristic	$O(n^2)$	**17.8%**
Bi-Section	$O(n^3)$ **[117]**	**13.56%**
Greedy	$O(n^2)$ **[118]**	**11.3%**
Genetic	$O(n^2)$ **[119]**	**3.98**

5.6 Results and Discussion

In this section, the performance of the proposed NSs is evaluated in terms of energy consumption and average piece delay.

Figure 5.9 shows the hourly variation of the number of active NSs. The number of active NSs increases considerably during peak hours of the day, which is expected. Interestingly, the number of active NSs does not follow the variation of the total traffic because the number also depends upon the spatial distribution of the traffic. Thus, at some non-peak hours, a greater number of active NSs may be needed to serve the traffic demand. In each hour, a set of installed NSs remained switched off resulting in 74% reduction in daily energy consumption.

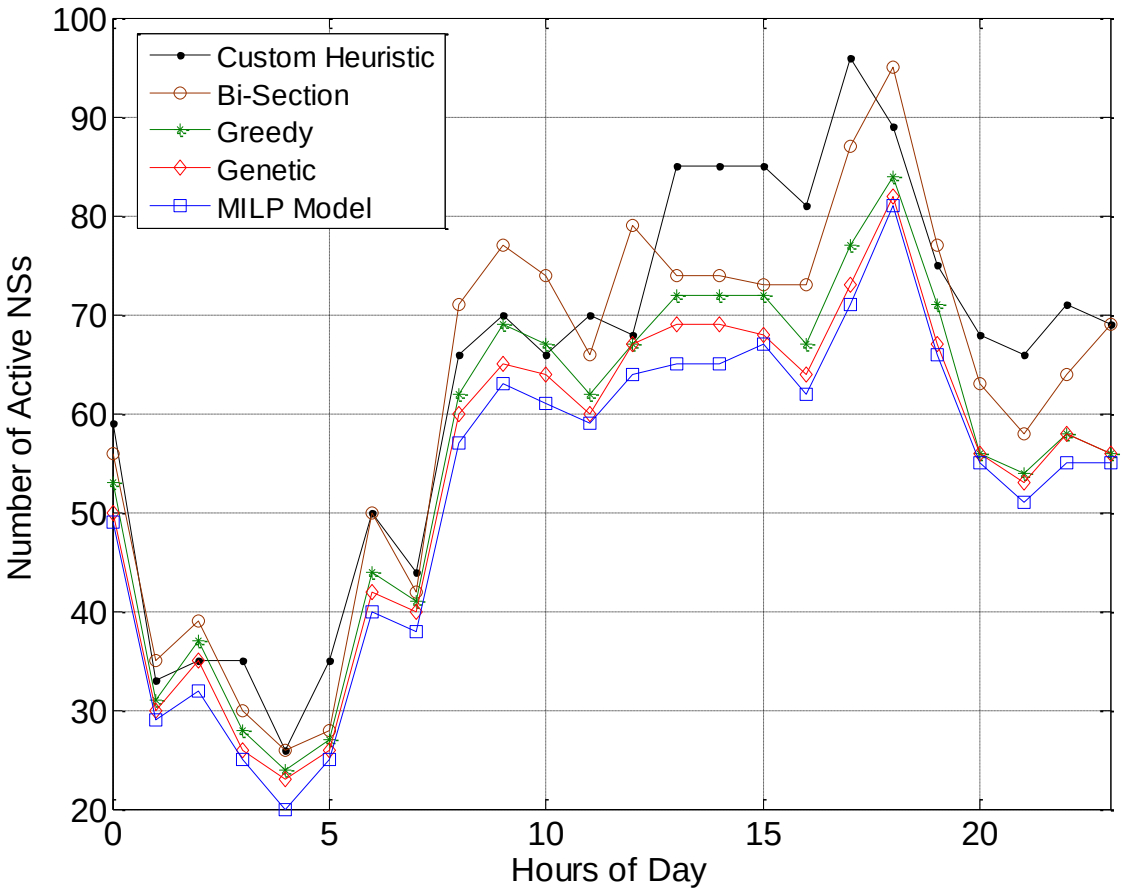


Figure 5.9 : Hourly variation of the number of active NSs with different optimisation methods.

Figure 5.10 shows the location of Traffic Points (TPs) which designate the traffic centroids that account for the traffic (IoT content/data) which the nearby vehicles request. Figure 5.11, shows the locations of active NSs at different hours of the day along with the installed NSs across the city roads (in the city area shown in Figure 5.1). By installed NSs, we mean those CS where NSs need to be active at least once in the whole day. The locations of the installed NSs can be obtained by the union of the hourly sets of active NSs over the whole day. The corresponding set of installed NSs is shown in figure 5.12. At each hour, a subset of the NSs is

active among the entire set of installed NSs, while the others are switched off. The MILP Model 1 reduces the required number of installed NSs from 398 to 202 resulting in 49% reduction in overall energy consumption of the NSs.

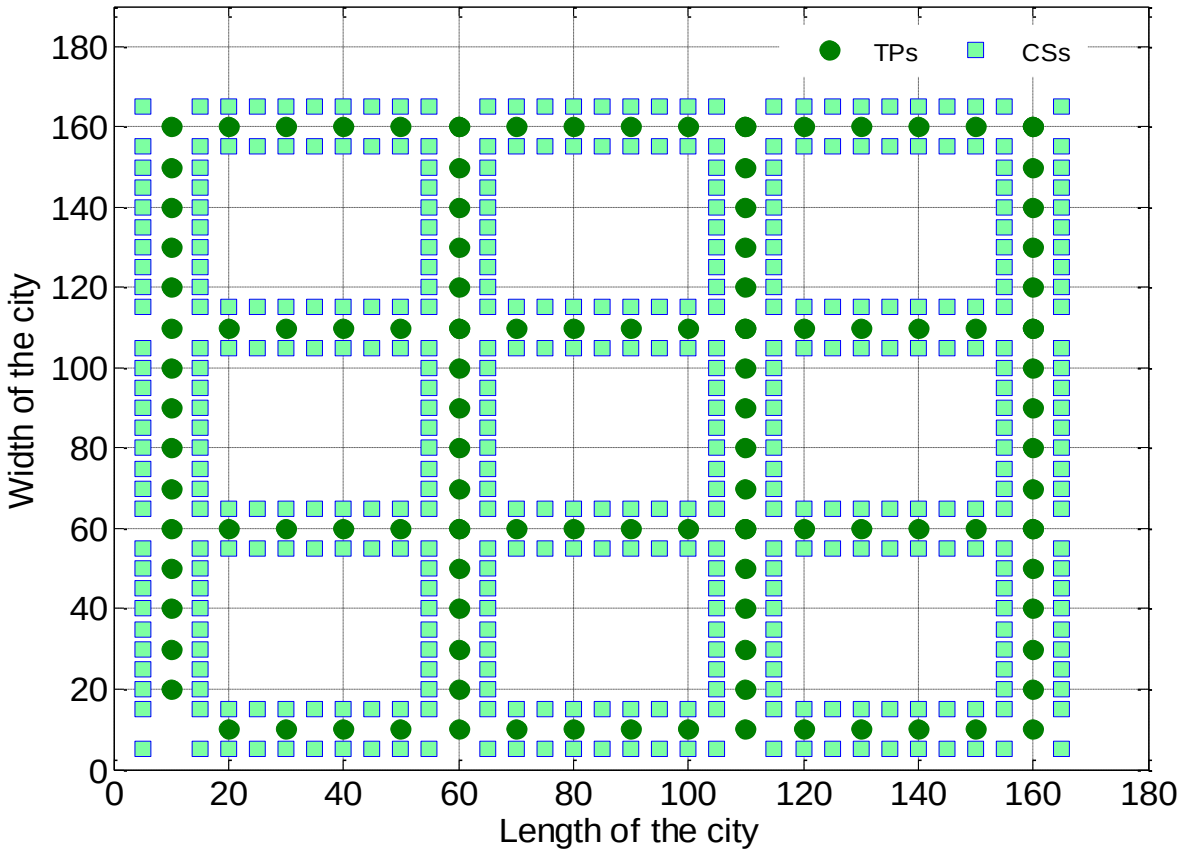


Figure 5.10: Location of the traffic points and candidate sites in studied city scenario.

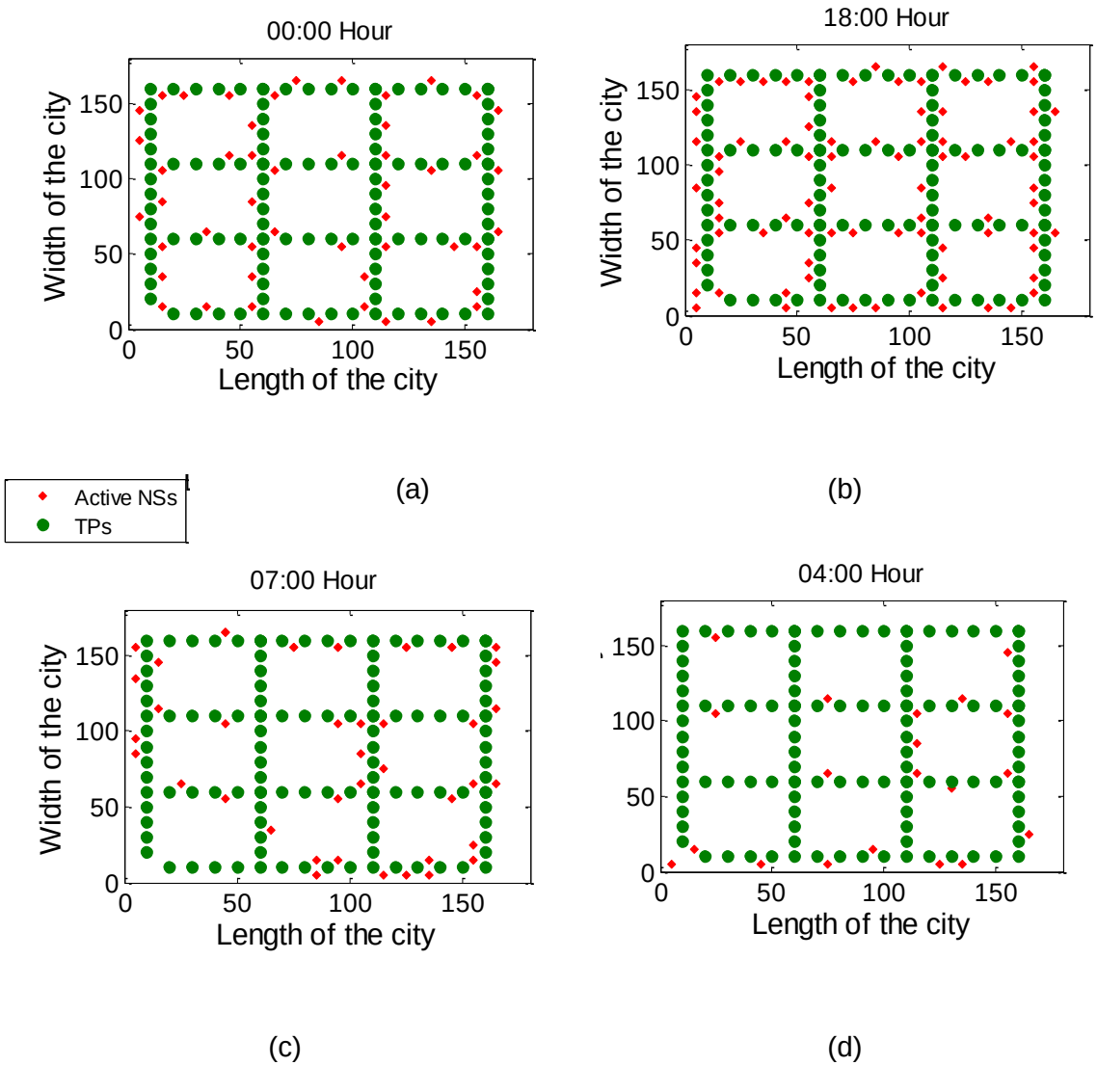


Figure 5.11: Location of the active NSs in smart city scenario for (a) Hour 00:00 (b)Hour 18:00 (c) Hour 07:00 and (d) Hour 04:00.

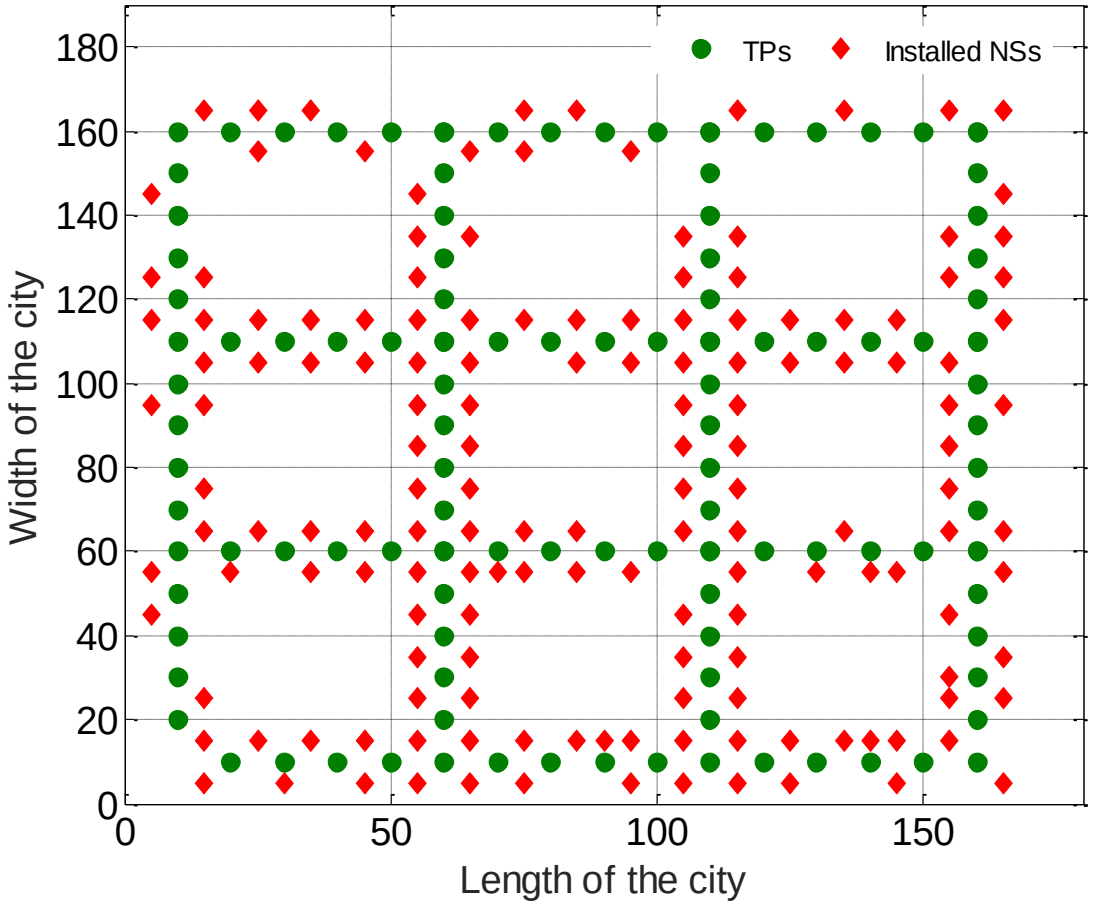


Figure 5.12: Location of the installed NSs in smart city scenario.

5.7 Conclusions

The number of active caches follows the hourly traffic variations, which is expected. Further, it is observed that the results obtained with MILP model and heuristic algorithm are in good agreement, with the exceptions of busy hours. This can be explained as follows. Since the MILP model possesses global knowledge of the traffic behaviour and restricts the number of active nano servers at each analysed time stretch (600 s), it adapts with greater degree of accuracy. Whereas, the heuristic algorithms operate with the instantaneous knowledge available, they do not restrict installing additional servers. Thus, a relatively higher number of installed cache points is experienced in case of using heuristic algorithm.The, genetic algorithm results have the least devition compared with the MILP.

Chapter 6: Transmission Energy Minimisation and impact of delay

6.1 Introduction

Power minimisation in the previous chapter was based on NSs locations and number optimisation. In this chapter we define two different techniques called (a) random sleep cycles (SC) and (b) energy aware rate adaptation (RA) for transmission energy minimisation of a NS. These give rise to reduced transmission power consumption at the cost of increasing average piece delay. In the remaining parts of this chapter, we define the corresponding analytic models and validate them with simulation.

6.2 Analytic model of a Nano Server with Random Sleep Cycles

In this model, each NS is considered to have a large buffer to hold the IoT piece requests and the arrival of IoT piece requests is a Poisson process. Since the NS operates multiple random sleep cycles [120] to save transmission energy, the NS is modelled as an M/D/1/∞ queue with queue length dependent vacations. The sleep durations of random sleep cycles are negative exponential

distributed with a certain mean value. Each sleep cycle (SC) is associated with a wake up overhead (E_{wo}) [120]. The total energy overhead of a NS with multiple sleep cycles is dependent upon the number of times (NO^s) the NS sleeps and wakes up within a given time duration.

We solve the queuing model with residual life approach as illustrated in Figure 6.1. The service duration of the NS is deterministic with mean, $\overline{X} = \mu^{-1}$ and variance $Var\ (X) = 0$. Further, the sleep cycles are Negative exponentially distributed with mean $\overline{S}$. The service durations and sleep cycle durations are independent and identically distributed (iid) random variables which are also independent of each other. The formulation sets, parameters and variables for the all the MILP models are defined in Table 6.1.

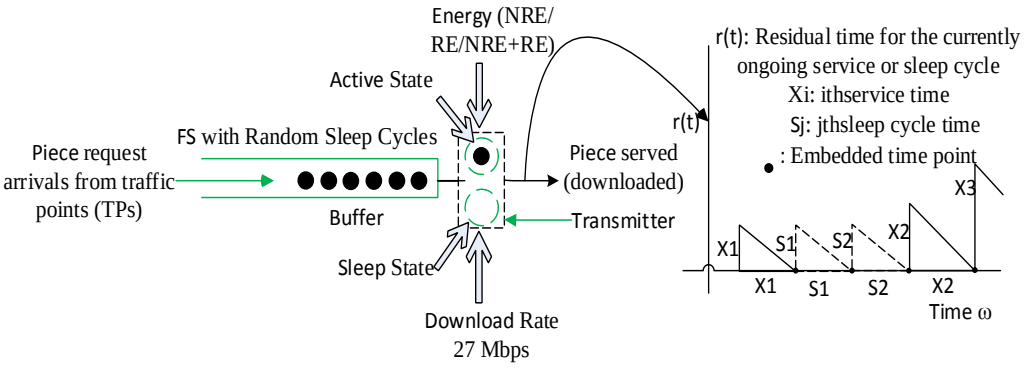


Figure 6.1: M/D/1/∞ queues with sleep cycles to represent a NS operation.

Table 6.1: List of Symbols.

Set	**Description**
TP	Set of traffic points
CS	Set of candidate sites
NS	Set of installed Fog computing Nano servers
$N[j]$	Set of neighbouring TPs for $NS\ n$
$N[n]$	Set of neighbouring NSs for $TPs\ j$
T	Set of time points within one hour (600 s each)
MILP Parameter	**Description**
$Bmax$	Maximum capacity of a NS (27 Mbps)
$Ndmax$	Maximum number of simultaneous downloads from a NS (9)
$drmax_{nt}$	Maximum data rate at $NS\ s$ at time t (i.e. 27 Mbps)
$Dmax_n$	Maximum acceptable average piece delay at $NS\ s$ at time t (i.e.13 s)
$Pmax_{nt}$	Maximum power consumption of at $NS\ s$ at time t (30 W [100])
$Pidle_{nt}$	Operational power consumption of $NS\ s$at time t 23 W [121]
$Ptxmax_{nt}$	Maximum transmission power consumption of a NS at time t given by $Pmax_{nt} - Pidle_{nt} = 7\ W$ [120]
Pw	Hourly available wind power

$Pmin_{nt}$	Minimum required power by $NS\ n$ at time t
$Esmax_n^s$	Maximum energy saving of $NS\ n$ with random sleep cycles
$Psmax_n^s$	Maximum power saving of $NS\ n$ with random sleep cycles
Ewo	Wake-up overhead energy in Joule
λ_{jt}	Traffic demand at $TP\ j$ at time t
$\overline{Smax}_n$	Maximum sleep duration at $NS\ n$
A	A constant, set to 600
σ	A constant, set to 10
MILP Variables	**Description**
Ptx_{nt}	Adaptive transmission power of $NS\ n$ at time t
$Ptxmin_{nt}$	Minimum transmission power of $NS\ n$ at time t according maximum acceptable delay
Es_n	Energy saving of $NS\ n$
Es_n^s	Energy saving of $NS\ n$ with random sleep cycles
Es_n^{ra}	Energy saving of $NS\ n$ with rate adaptation
Ps_n^s	Power saving of $NS\ n\ t$ with random sleep cycles
Ps_n^{ra}	Power saving of $NS\ n$ with rate adaptation
α_n	Equals 1 if $NS\ n$ is on, equals 0 otherwise
λ_{njt}	Traffic between $NS\ n$ and TP j at time t
λ_n	Traffic demand at $NS\ n$ time t
δ_{njt}	Equals 1 if $NS\ n$ is transmitting IoT content to $TP\ j$, equals

	0 otherwise
Pre_{nt}	The part of the available renewable power used by $NS\ n$ at time t
$Dmin_n$	Minimum average piece delay at $NS\ s$ at time t(when data rate is $drmax_{nt}$
dr_{nt}	Adaptive data rate at $NS\ n$ at time t
$\bar{S}_n$	Sleep duration at $NS\ n$
$\overline{Smax}_n$	Maximum Sleep duration at $NS\ n$
NOs_n	Number of sleep cycles at $NS\ n$
Dra_n	Average piece delay at the rate adaptive $NS\ n$
Dsc_n	Average piece delay at $NS\ n$ with sleep cycles
Index	**Description**
n	Index of Nano Servers (NS)
j	Index of Traffic Point (TP)
t	Index of Time Point (T)

Let $r(\omega)$ be the residual time for the ongoing service or sleep. Time average of $r(\omega)$, $(0 \leq \omega \leq t)$ can be given as

$$\overline{R}_t = \frac{1}{t}\int_0^t r(\omega)\, d\omega = \frac{1}{t}\sum_{i=1}^{N_\lambda(t)} \frac{1}{2} X_i^2 + \frac{1}{t}\sum_{j=1}^{NOs(t)} \frac{1}{2} S_j^2 \quad (6.1)$$

where $N_\lambda(t)$ is the number of IoT piece requests that arrive within $(0, t)$, and ω is the instantaneous time. $NOs(t)$ is the number of sleep cycles (sleep count) within $(0, t)$.

Therefore, the average residual time is given as

$$\overline{R} = \lim_{t\to\infty} \overline{R}_t = \lim_{t\to\infty} \left[\frac{1}{t}\sum_{i=1}^{N_\lambda(t)} \frac{1}{2}X_i^2 + \frac{1}{t}\sum_{j=1}^{NOs(t)} \frac{1}{2}S_j^2\right] \tag{6.2}$$

$$\overline{R} = \lim_{t\to\infty}\left[\frac{N_\lambda(t)}{t}\frac{\sum_{i=1}^{N_\lambda(t)}\frac{1}{2}X_i^2}{N_\lambda(t)}\right] + \lim_{t\to\infty}\left[\frac{N^s(t)}{t}\frac{\sum_{j=1}^{NOs(t)}\frac{1}{2}S_j^2}{NOs(t)}\right] \tag{6.3}$$

$$\overline{R} = \frac{1}{2}\lambda\overline{X^2} + \frac{1}{2}\lim_{t\to\infty}\frac{1}{\frac{t(1-\rho)}{NOs(t)}}\overline{S^2}(1-\rho) \tag{6.4}$$

$$\overline{R} = \frac{1}{2}\lambda\overline{X^2} + \frac{1}{2}(1-\rho)\frac{\overline{S^2}}{\bar{S}} \tag{6.5}$$

where the offered load ρ also corresponds to system utilisation (U_s), is defined as

$$\rho = U_s = \lambda\bar{X} \tag{6.6}$$

The relation between ρ, NOs and $\bar{S}$ can be defined as

$$\frac{t(1-\rho)}{NOs} = \bar{S} \tag{6.7}$$

where t represents a time duration.

The average waiting time for an IoT piece at the NS can be computed using Little's theorem [122] as

$$Wq = NOq\ \bar{X} + \overline{R} \tag{6.8}$$

$$Wq\ = \lambda \mathrm{W_q} \bar{X} + \overline{R} \tag{6.9}$$

where NOq represents the number of IoT piece requests waiting in the queue.

Therefore,

$$Wq = \frac{\overline{R}}{(1-\rho)} \tag{6.10}$$

Since $= \lambda \bar{X}$, using Equation (6.2) and Equation (6.10) we get

$$Wq = \frac{\lambda \overline{X^2}}{2(1-\rho)} + \frac{\overline{S^2}}{2\bar{S}} \tag{6.11}$$

The average delay of an IoT piece (Dsc) including the waiting delay (Wq) and the service time ($\overline{X}$) can be computed as

$$Dsc = Wq + \bar{X} \tag{6.12}$$

$$Dsc = \bar{X} + \frac{\lambda \overline{X^2}}{2(1-\rho)} + \frac{\overline{S^2}}{2\bar{S}} \tag{6.13}$$

since $\bar{X} = \frac{1}{dr}$, $\overline{X^2} = Var(X) + (\bar{X})^2 = \frac{1}{dr^2}$ ($Var(X) = 0$ as data rate is determistic)

$$Dsc = \frac{1}{dr} + \frac{\lambda}{2dr^2(1-\rho)} + \frac{\overline{S^2}}{2\bar{S}} \tag{6.14}$$

since $\overline{S^2} = Var(S) + (\bar{S})^2 = (\bar{S})^2 + (\bar{S})^2 = 2(\bar{S})^2$ (6.15)

$$Dsc = \frac{1}{dr} + \frac{\rho}{2\mu(1-\rho)} + \bar{S} \tag{6.16}$$

The average number of pieces in the system ($Nsys$) from Little's theorem [122] can be written as

$$Nsys = \lambda Dsc \tag{6.17}$$

$$Nsys = \frac{\lambda}{dr} + \frac{\rho\lambda}{2\mu(1-\rho)} + \bar{S} \tag{6.18}$$

$$Nsys = \rho + \frac{\rho^2}{2(1-\rho)} + \bar{S} \tag{6.19}$$

The energy savings (Es) per hour through sleep cycles for the NS by using (6.7) can be obtained as

$$Es = (1 - U_s) \times PT \times T - (Ewo \times NOs) \tag{6.20}$$

$$Es = (1 - \rho) \times PT \times T - (Ewo \times NOs) \tag{6.21}$$

$$Es = \frac{NOs \times \bar{S}}{T} \times PT \times T - (Ewo \times NOs) \tag{6.22}$$

$$Es = NOs \times (\bar{S}\, PT - Ewo) \tag{6.23}$$

where U_s is the system utilisation, PT is the NS's transmitter circuitry power consumption and Ewo is the NS's wake-up overhead. The number of sleep cycles i.e. sleep count (N^s) per unit time can be computed from (6.4) as

$$\lim_{t\to\infty}\frac{NOs(t)}{t} = NOs = \frac{(1-\rho)}{\bar{S}} \tag{6.24}$$

Rearranging (6.8), in terms of $\bar{S}$, we obtain

$$\bar{S} = Dsc - \frac{2dr-\lambda}{2dr(dr-\lambda)} \tag{6.25}$$

The MILP constraint for transmission energy minimisation through sleep cycles can therefore be written as:

If dr is selected as $dr = drmax$, then service time is short and sleep duration is maximised as a result the power saving is maximum and delay $Dsc = Dmax$. Therefore the motivation for maximum power saving at NS can be obtained from (6.25) as

$$\overline{Smax}_n = Dmax_n - \frac{2dr-\lambda_n}{2drmax(drmax-\lambda_n)} \tag{6.26}$$

$$\forall\, n \in NS$$

The sleep count at $NS\ n$ is

$$NOs_n = \left(1-\frac{\lambda_n}{drmax}\right)1/\overline{Smax}_n \tag{6.27}$$

$$\forall\, n \in NS$$

Therefore, from (6.23) the power saving is given by

$$Esmax_n = NOs(\overline{Smax}_n Ptx_{nt} - Ewo) \quad (6.28)$$

$$\forall\, n \,\epsilon\, NS$$

6.3 Analytic Model of a Nano Server with Rate Adaptation

For rate adaptive NSs, the data rate of a NS can adapt according to the transmission energy for that NS. Assuming that the data rate is linearly proportional to the transmission energy [110], the adaptive data rate of a NS (dr) can be expressed in terms of transmission energy (PT) as

$$dr = ({}^{Ptx}/_{Ptxmax})drmax \quad (6.29)$$

where $drmax$ is the non adaptive (maximum) data rate and $Ptxmax$ is the maximum transmitting power (e.g. 7 W). Invariably, $Ptxmin \leq Ptx \leq Ptxmax$ and $drmin \leq dr \leq drmax$. Thus, the transmission energy consumption of a NS reduces compared to that of a non-rate adaptive NS. This, in turn, further reduces the total energy consumption.

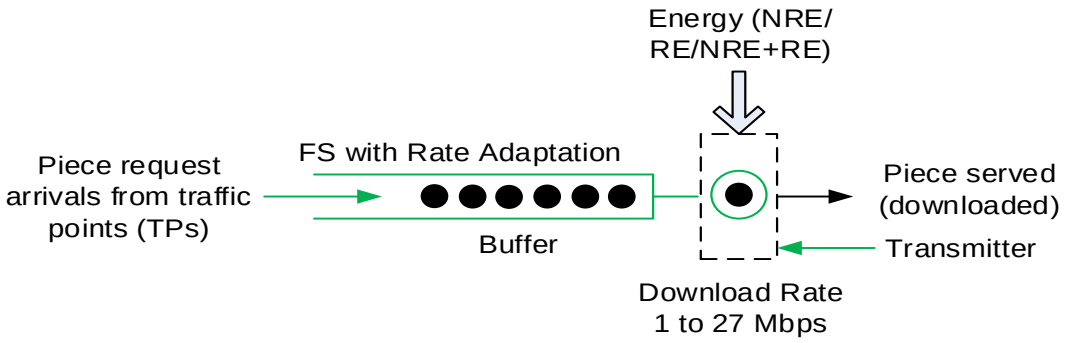


Figure 6.2: M/D/1/∞ queues with rate adaptation to represent a NS operation.

To find out a relationship between data rate at each NS and the average IoT piece delay, we analyse the stochastic properties of arrival and departure processes of a NS. The arrival process of IoT piece requests at any NS is random, therefore can be modelled as a Poison process. The IoT piece size is fixed (deterministic) and the data rate is also fixed for a fixed supply of transmission energy (does not change in a short time). Therefore, the service time of an IoT piece is deterministic. Such a NS can be modelled as a $M/D/1$ queue (Model 6.2). To find out the expression for average IoT piece delay in a rate adaptive NS, we set $\bar{S} = 0$ in (6.8) and rearrange as

$$Dra = (2\,dr - \lambda)\,{}^{1}\!/_{2(dr)^2 - 2dr\lambda} \tag{6.30}$$

The MILP constraint for transmission energy minimisation through rate adaptation can therefore be written as: when $Dra = Dmax$, $dr = drmin$, which is the condition for maximum energy savings through rate adaptation at a NS. Solving (6.30), we obtained the minimum data rate for $NS\ n$ as

$$drmin_{nt} = \frac{2\lambda_n Dmax_n + 2 \pm \sqrt{4\lambda_n^2 (Dmax_n)^2 + 4}}{4Dmax_n} \tag{6.31}$$

$$\forall\ n\ \epsilon\ NS, \forall\ t\epsilon T$$

The minimum power consumption of a rate-adaptive NS can be expressed as

$$Pmin_{nt} = Ptxmin_{nt} + Pidle_{nt} \tag{6.32}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

where $Pmin_{nt}$ is the minimum power required by NS n for serving its traffic with the maximum acceptable IoT piece delay. The quantity $Pmin_{nt}$ is computed by expressing (6.31) as:

when $$\sum_{t\epsilon T} \sum_{j\epsilon tp[n]} \delta_{njt} \neq 0 \tag{6.33}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$dr_{nt} = \frac{1}{2Dra_n} \sum_{t\epsilon T} \sum_{j\epsilon N[j]} \lambda_{njt}\ Dra_n + 1 \pm \sqrt{\left[\sum_{t\epsilon T} \sum_{j\epsilon N[j]} \lambda_{njt}\ Dra_n + 1\right]^2} \tag{6.34}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

For minimum adaptive data rate ($drmin_{nt}$), we write

$$Dra_n = Dmax_n = 13\ s \tag{6.33}$$

$$\forall\, n \,\epsilon\, NS$$

$$Ptxmin_{nt} = \frac{drmin_{nt}}{drmax_{nt}}\ Ptxmax_{nt} \tag{6.34}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

When the energy source is non-renewable, the data rate of a rate adaptive NS reaches the minimum to maintain the average piece delay below the acceptable limit (13 s).

6.4 Model Validation of a NS with Sleep Cycles and Rate Adaptation

To validate the models, we obtain the energy savings from 6.2 and 6.3 for a fixed average piece delay through analytic modelling and simulation. For analytic results, the corresponding expressions for energy savings are given below.

From Model 6.1, we obtain a relationship between energy savings (Es_{nt}^{s}) and mean arrival rate of piece requests (λ) in terms of fixed average piece delay ($D = 5\ s$) using (6.23), (6.27) and (6.28) as

$$Es_n^s = \quad Ptxmax_{nt}\left(1 - \frac{\lambda}{drmax_{nt}}\right)T -$$

$$(drmax_{nt} - \lambda)^2 Ewo\left(\frac{1}{2drmax_{nt}{}^2 Dsc_n - 2drmax_{nt}(Dsc_n\lambda + 1) + \lambda}\right) \tag{6.35}$$

$$\forall\ n\ \epsilon\ NS, \forall\ t\epsilon T$$

Similarly from Model 6.2, we obtain a relationship between energy savings (E_s) and mean arrival rate of piece requests (λ) in terms of fixed average piece delay ($D = 5s$) using (6.29), (6.30) and (6.34) as

$$Es_n^{ra} = Ptxmax_{nt}\left(1 - \frac{Dmin_n(2\lambda Dra_n + 2 \pm \sqrt{4\lambda^2 Dra_n{}^2 + 4})}{D_2 \times (2\lambda Dmin_n + 2 \pm \sqrt{4\lambda^2 Dmin_n{}^2 + 4})}\right) \tag{6.36}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

For simulation results, we develop respective be-spoke Java programs using the flowchart shown in figure 6.3 for rate adaptation and for sleep cycle we used the simulation described in Figure 3.4 in Chapter 3.

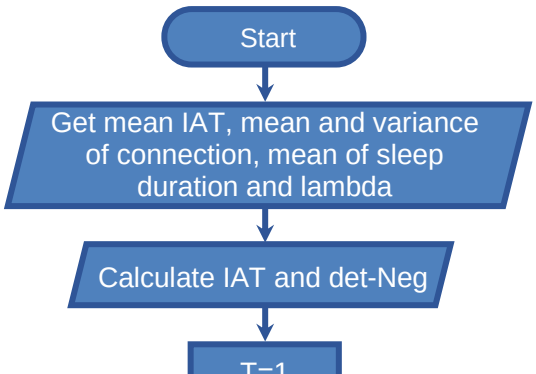
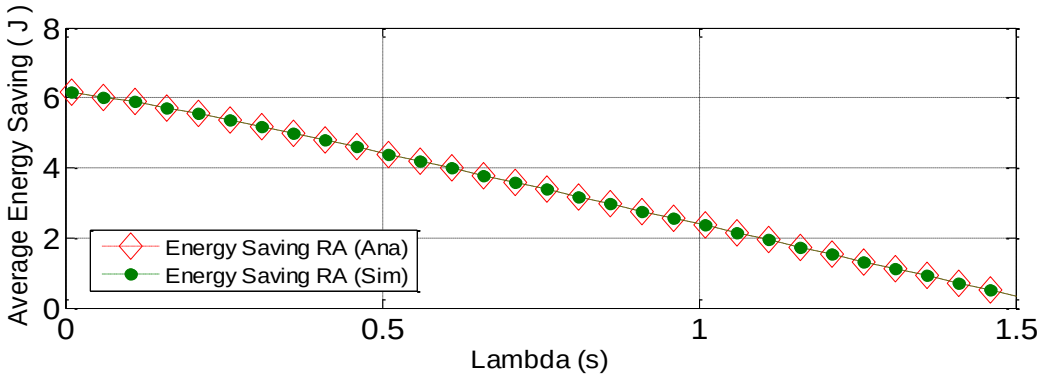


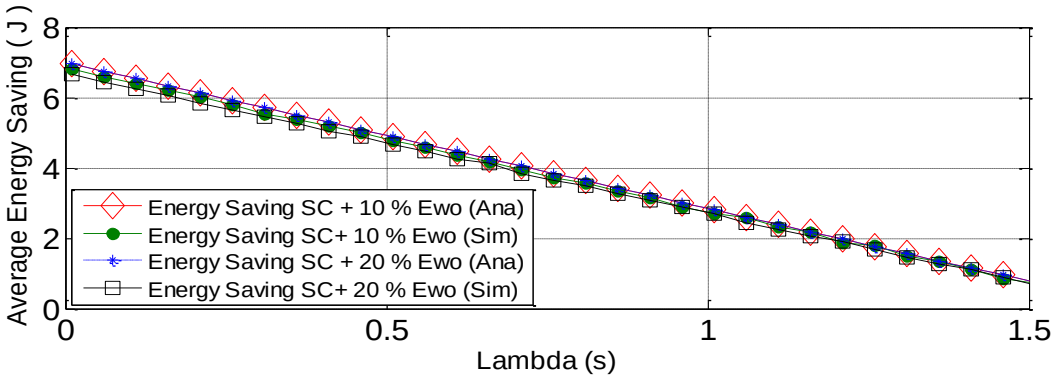
Figure 6.3 : Delay and Number of sleep with a rate adaptive nano Server.

Figure 6.4 shows the average energy savings achieved by Rate adaptive NS (a) and NS with sleep cycles (b) for a NS with varying arrival of piece requests (lambda) when the respective average piece delay is maintained at 5 s. As can

be seen in Figure 6.1, the analytic and simulation results closely agree with each other, which serve to validate the proposed models. Since there is no reference wake-up overhead value for a NS available in the literature, we assume two cases where the wake-up process consumes 10% and 20% of the maximum transmission power of a NS. It is observed in Figure 6.4(a) that the energy saving decreases with increasing load (lambda). With the increase in load, the sleep duration need to be reduced in order to maintain the fixed average piece delay ($5\ s$). These, in turn, decrease the energy savings. It is evident that the higher wake-up overhead (20%) results in lower energy savings for the same load. Figure 6.3 (b) shows a similar trend as that of Figure 6.4 (a). The wake up energy and the energy saving don't have any relationship as Figure 6.4 (b) shows the wake up energy doesn't have any effect on energy saving. With rate adaptation, the data rate of a NS increases with increase in load. This results in decrease in energy savings. Figure 6.5 (a) and 6.5 (b) show the average piece delay for varying load. Since the delay is kept constant at $5\ s$, we only find minor variation in the simulation which is acceptable due to the inherent randomness.



(a)

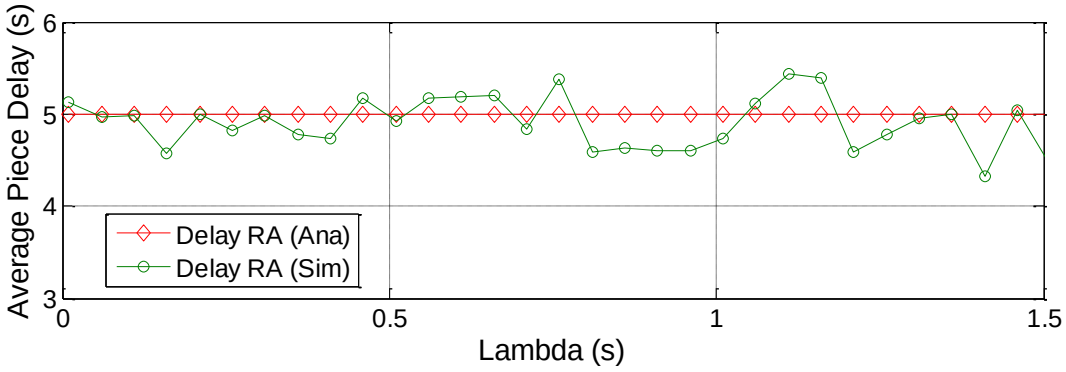


(b)

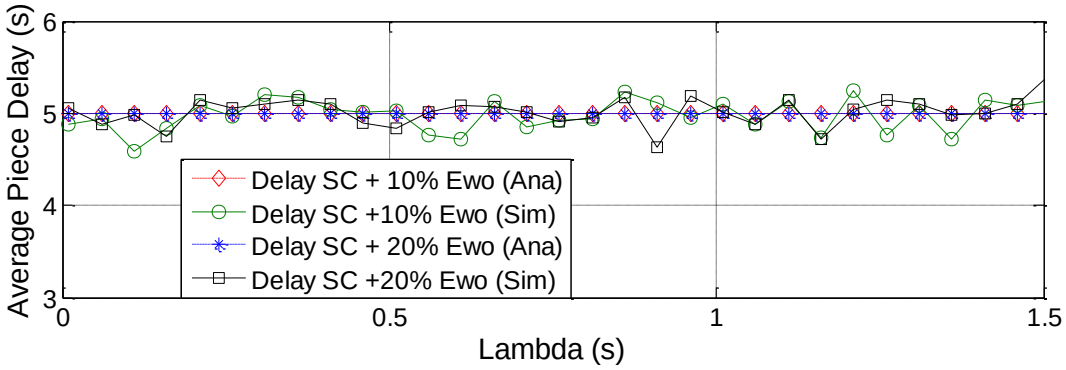
Figure 6.4: Average energy saving achieved by a typical Nano Server (a) rate adaptation (b) sleep cycle.

Figure 6.6 shows the effect on average piece delay with varying amount of energy savings for a moderate loaded NS for Model 6.1 and Model 6.2. As expected, in both models, higher energy savings can be achieved at the cost of higher average piece delay. At the given energy saving, if the wake up overhead energy is larger, then power available for transmission is smaller so delay is higher. It is interesting to note that in lower energy savings region, Model 6.2 performs better than Model 6.1 and vice versa. Further to note that

since in rate adaptation energy savings is directly related to the data rate, the rate of increase in average piece delay is higher than that of sleep cycles model.



(a)



(b)

Figure 6.5: Average piece delay at a typical NS (a) rate adaptation (b) sleep cycle.

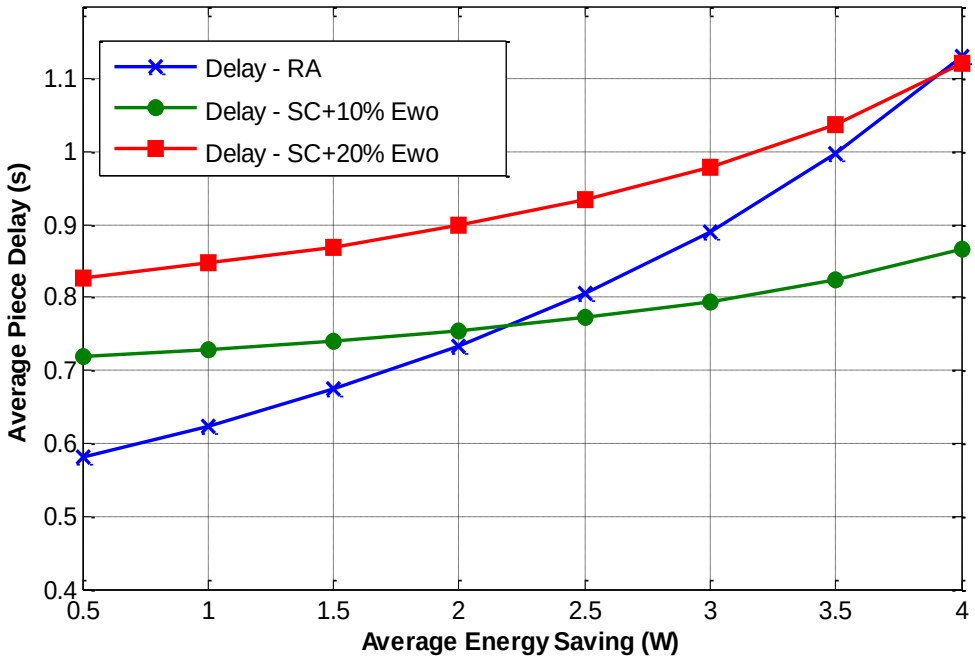


Figure 6.6 : Relationship between average piece delay and average energy savings.

6.5 MILP Model for Optimum usage of Non-Renewable Energy and Renewable Energy Mix

We utilise the renewable (wind) energy (RE) available from wind farms to reduce the non-renewable energy consumption of the NSs. Installation of a wind turbine at each NS is aesthetically not suitable for a city, so for this chapter we used a wind farm. The corresponding MILP models (6.3 and 6.4) minimise the non-renewable energy (NRE) consumption subject to the average piece delay constraint and subsequently improve the delay with available RE. The models ensure that heavily loaded NSs would receive greater RE to improve

overall user experience. In MILP model 6.3 uses NSs with sleep cycles and MILP model 5 uses rate adaptive NSs.

6.5.1 Sleep Cycle

Model 6.3 is developed based on Location optimisation MILP Model in Chapter 5 to reduce complexity of the MILP. The optimised NSs location is used as input here therefore the objective function of the MILP Model 6.3 is to minimise power consumption by maximising energy saving ($\forall t \epsilon T, \alpha_n = 1$). NOs_n is not a variable in this MILP, so the number of sleep depends on packet arrival thus, the average piece request (λ) needs to be determined which is calculated for the location optimisation model. The objective function of the MILP model is

$$\boldsymbol{Minimise}$$

$$\sum_{t\epsilon T}\sum_{n\epsilon NS} \alpha_n(Ptxmax_{nt} + Pidle_{nt} - Ps_n^s - Pre_{nt}) \tag{6.37}$$

$$\forall\, n \epsilon NS, \forall\, t\epsilon T$$

where

$$Ps_n^s = {Es_n^s}/{T} \tag{6.38}$$

The constraints of the MILP model are given below.

$$\overline{S_n} \leq \overline{Smax}_n \tag{6.39}$$

$$\forall\, n \epsilon NS$$

$$Pre_{nt} \leq Ptxmax_{nt} + Pidle_{nt} - Psmax_n^s \quad (6.40)$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

where from (6.28)

$$Psmax_n^s = NOs(\overline{Smax}_n Ptx_{nt} - Ewo)/T \quad (6.41)$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$\bar{S}_n = Dsc_n - \frac{2drmax_{nt} - \lambda_n}{2drmax_{nt}(drmax_{nt} - \lambda_n)} \quad (6.42)$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$\lambda_n = \sum_{t\epsilon T} \sum_{j\epsilon N[j]} \lambda_{njt} \quad (6.43)$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$\bar{S}_k - \bar{S}_i \geq \sigma\,(\lambda_i - \lambda_k) \quad (6.44)$$

$$\forall\, i \,\&\, k \,\epsilon\, NS, i \neq k$$

Equations (6.40) ensures that the renewable energy at $NS\ n$ is less than or equal to the required power consumption of n. The sleep duration constraint is as Equation (6.42).The sleep duration $\bar{S}_n$ of a NS is tunable based on the available renewable energy. The demand at each NS is the computed λ_n in location optimisation MILP model in Chapter 5. The demand at each *NS* is computed as Equation (6.43).

With excess RE, the MILP model 6.3 reduces sleep duration at busy NSs which experience higher piece delay. This is achieved by Equation (6.44) where σ is a

constant which makes the right hand side (traffic demand) comparable with the left hand side (sleep duration).

6.5.2 Rate Adaptation

The objective function of the MILP model 6.4 is

$$\boldsymbol{Minimize}$$

$$\sum_{t\epsilon T}\sum_{n\epsilon NS} \alpha_n(Ptx_{nt} + Pidle_{nt} - Pre_{nt}) \tag{6.45}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

The constraints of the MILP model 6.4 apart from Equation (6.30) and (6.43), are the following additional constraints given below.

$$Dra_n \leq Dmax_n \tag{6.46}$$

$$\forall\, n \,\epsilon\, NS$$

$$Pre_{nt} \leq Ptx_{nt} + Pidle_{nt} \tag{6.47}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$Ptx_{nt} + Pidle_{nt} \geq Pmin_{nt} \tag{6.48}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$Ptx_{nt} \leq Ptxmax_{nt} \tag{6.49}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$dr_{nt} = drmax_{nt}\,\frac{Ptx_{nt}}{Ptxmax_{nt}} \tag{6.50}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$X_{nt} = \ Ptx_{nt} \ \alpha_n \tag{6.51}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$0 \leq X_{nt} \leq Ptxmax_{nt} \tag{6.52}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$X_{nt} \ \leq \ \alpha_n Ptxmax_{nt} \tag{6.53}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$X_{nt} \geq Ptx_{nt} - Ptxmax_{nt}(1 - \alpha_n) \tag{6.54}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$dr_{it} - \ dr_{kt} \ \geq \sigma\ (\lambda_i \ - \ \lambda_k) \tag{6.55}$$

$$\forall\, i \ \&\ \ k \,\epsilon\, NS, i \neq k$$

Equation (6.46) together with (6.30) is used to limit delay. The minimum required energy ($Pmin_{nt}$) is calculated using (6.18) where the data rate of a NS is adaptive based on the available renewable energy for transmission. Equation (6.40) is modified as Equation (6.47).The transmission energy constraint is defined as Equations (6.48) and (6.49).The constraint for adaptive data rate are given as Equation (6.50). Equation (6.50) calculates the adaptive data rate of each NS at each time point t.

To compute the adaptive transmission power of $NS\ n$ if it has a connection with any TP ($\alpha_n = 1$), linearisation is needed so X_{nt} defended as Equation (6.51).The following equations (6.52 - 6.54) are used to remove the non-linearity of multiplying two variables with linear relationship. With excess RE, the proposed model improves data rate at busy NSs which experience lower

data rate. This is achieved by Equation (6.55) where σ is a constant which make the right hand side (traffic demand) comparable with the left hand side (data rate).

6.6 Genetic Algorithms Validation

We develop another genetic algorithm (GA2) for validating Model 6.3 and 6.4, which determines the optimum mixture of renewable and non-renewable energy based on load profile. In GA2, we start working with a population of individuals (corresponds to a random amount of renewable energy for each TP) having a "fitness" (average piece delay) and then allow that population to evolve to a more fit state. The better-fit individuals having lower delay would be more likely to mate than an individual that is poorly fit to survive in the new environment. Thus, the genes of the lower delay solutions are more likely to be passed on to the offspring.

Figure 6.7, shows the average piece delay achieved at each hour for (i) the CPs with random sleep cycles with different wake up overhead energy and (ii) the rate adaptive CPs through MILP and Genetic Algorithms. The GA result deviates 5.9% from these of RA and 6.1% and 6.5% from these of SCs 10% & SCs 20%. The flowchart for genetic algorithm is shown in figure 6.8. This flowchart is quite similar to flowchart 5.4 apart from our initial population and fitness factor. Chromosomes have renewable power allocations as array elements and the size of each array is equal to the total number of NSs each

hour. Renewable power is converted to delay through equation (6.23) and (6.28) and also for rate adaptation equations (6.29) and (6.30) are used. Fitness is the total delay so chromosomes with the lowest delay are the fittest.

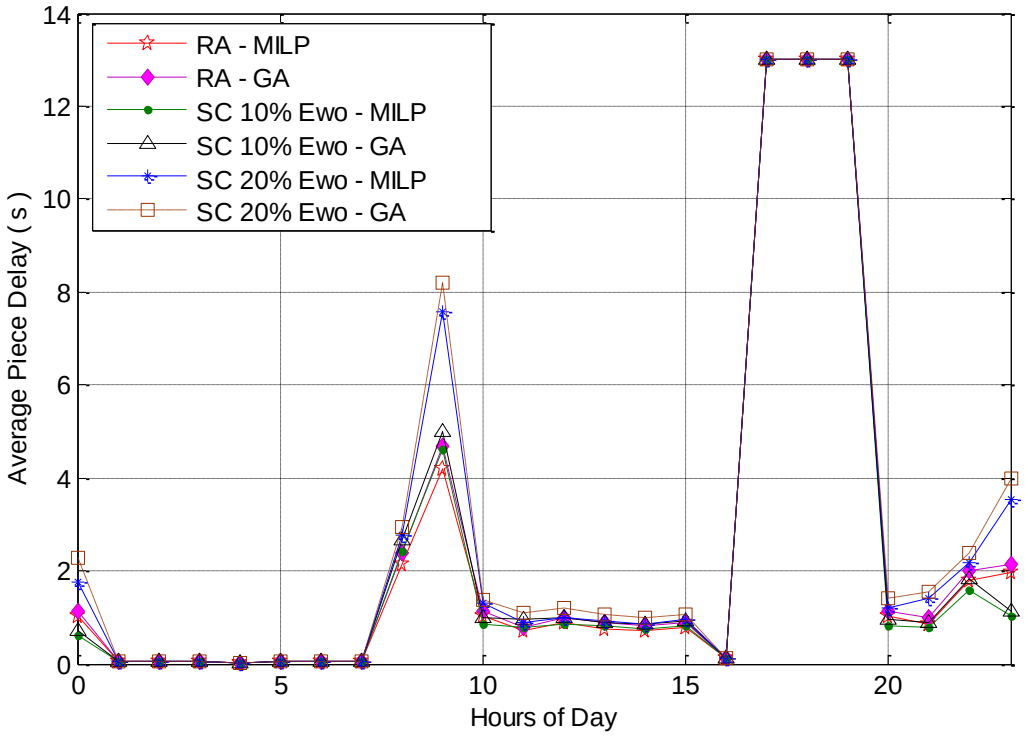


Figure 6.7: Average piece delay validation via genetic algorithm.

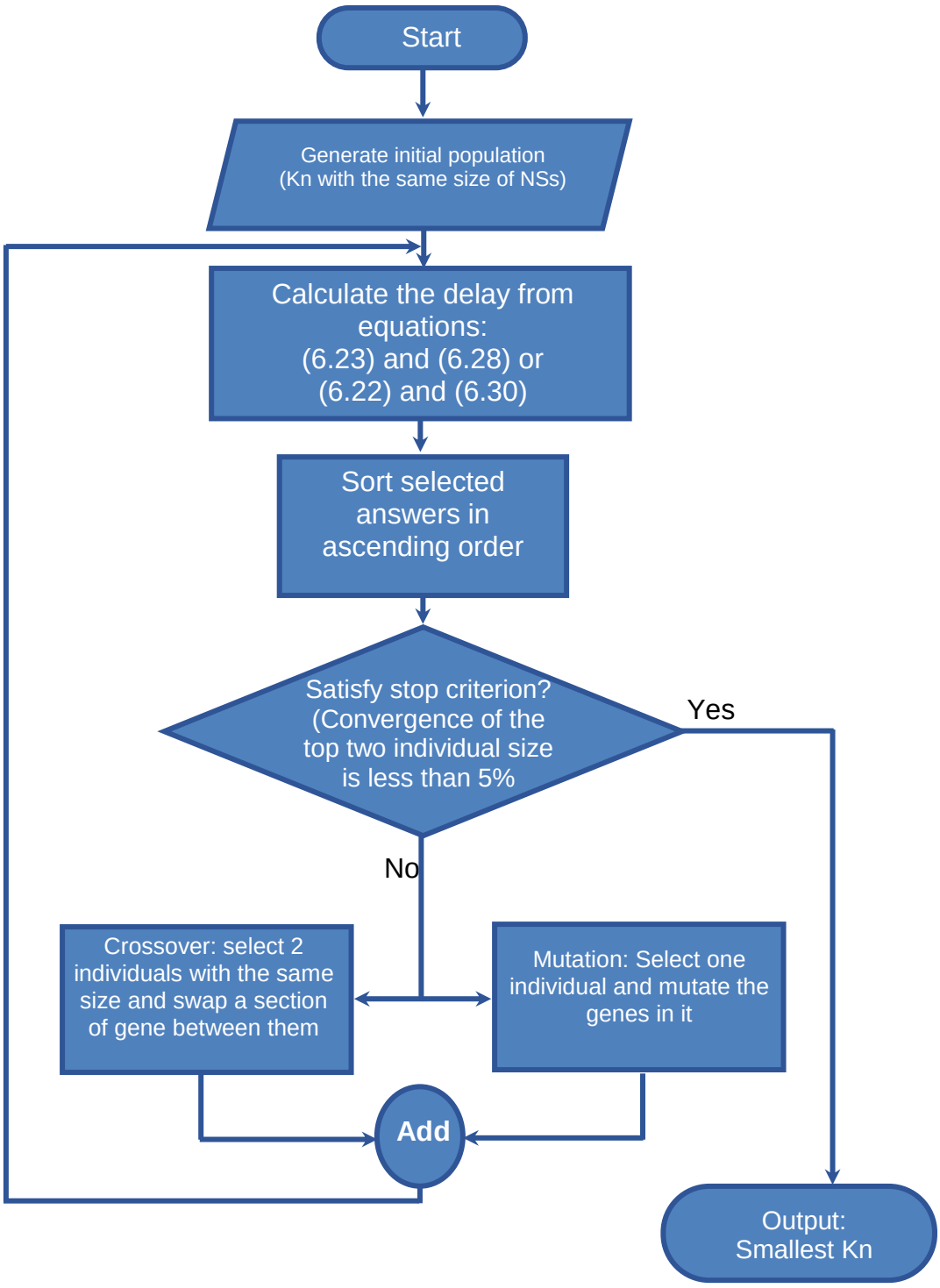


Figure 6.8: Optimum usage of Non-Renewable Energy and Renewable - Genetic.

6.7 Results and Discussion

In this chapter, the performance of the proposed NSs is evaluated in terms of energy consumption and average piece delay. Figure 6.9 shows the available wind energy from the wind farm, the minimum and the maximum required energies by the installed NSs. By minimum required energy, we mean the least amount required to maintain the average piece delay under the mentioned bound. The relationship is obtained through the queuing model in this chapter. At early hours of the day, the available wind energy is considerably higher compared to the required energies. However, it becomes comparable during peak hours due to higher traffic load at those periods. Since the available wind energy and the required energy (minimum or maximum) are independent of each other, the available energy is even lower than the minimum required energy at some instances i.e. hours 17-19. Consequently, these hours require NRE to meet the deficiency.

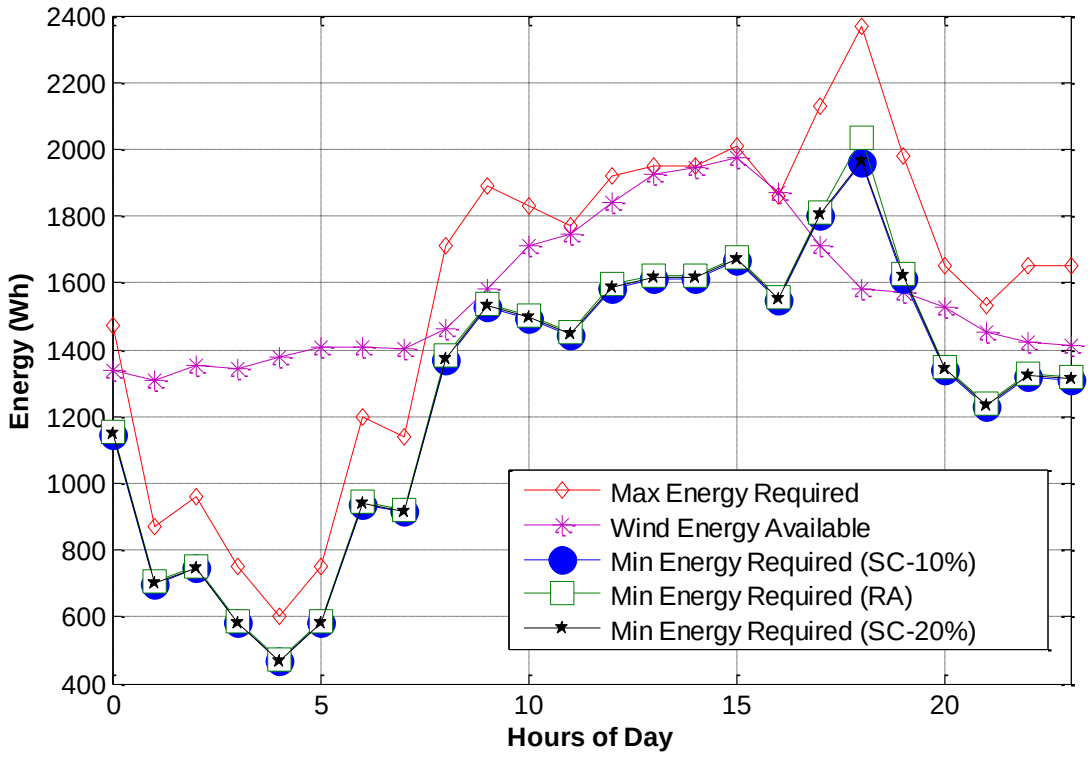


Figure 6.9 : Available and required energy for the rate-adaptive NSs.

The usage of the optimal energy consumption of the rate adaptive NSs and NSs with Sleep Cycles is shown in Figure 6.10 when an optimal mix of NRE and RE are used (through MILP Model 2 and 3). As can be seen that apart from a few occasions where the available wind energy was lower than that of the minimum required energy, the available wind energy on its own was sufficient to operate the NSs while improving the data rate.

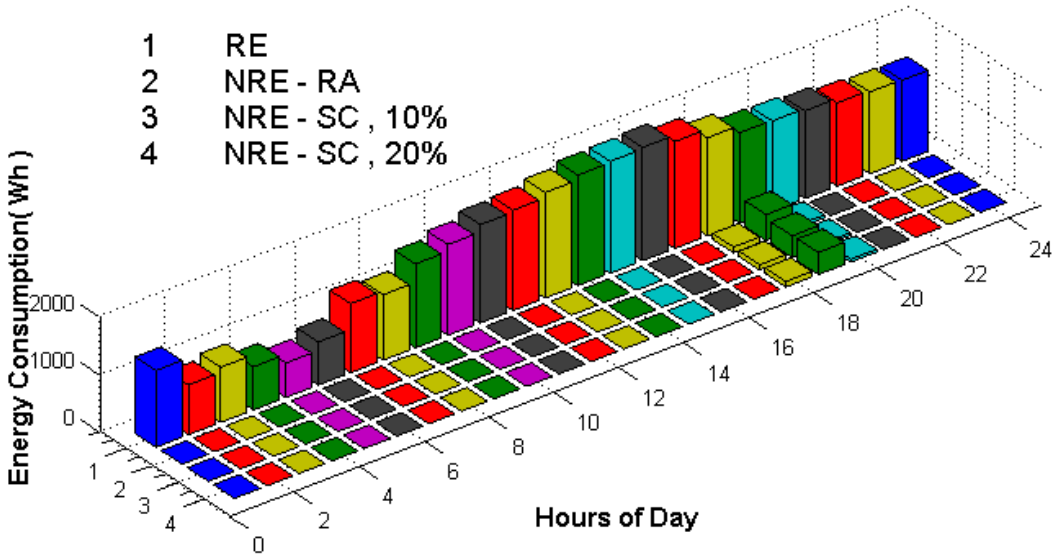
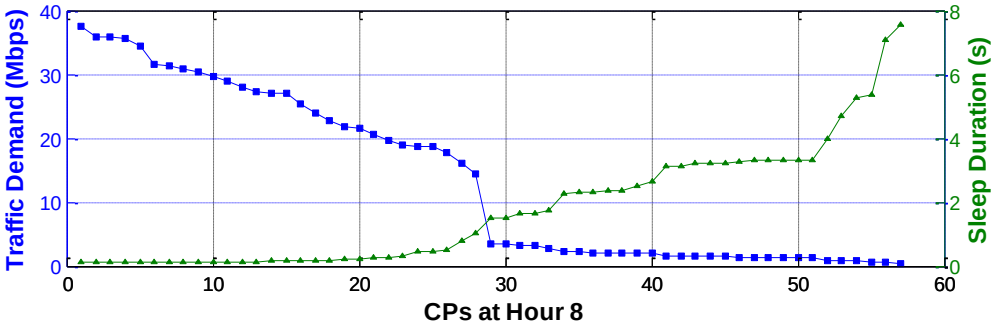


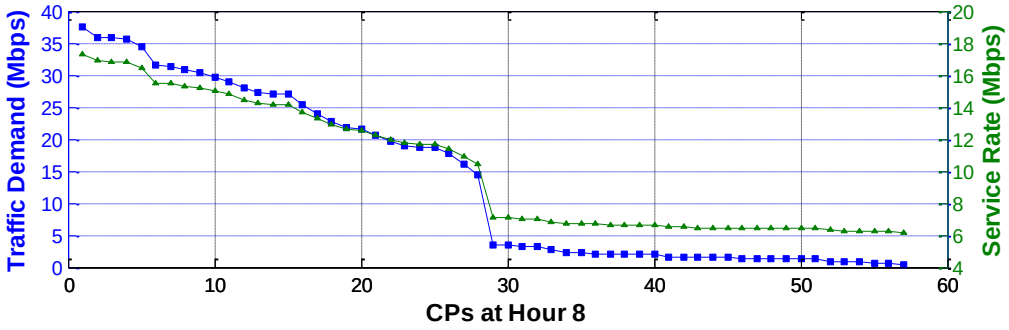
Figure 6.10: Optimal energy consumption of the adaptive NSs.

Chapter 5 has investigated the location of NSs that have to be active at each point in the day, in this chapter an hour of day selected, 08:00 a.m, at which the vehicular traffic is at normal level just before the 09:00 am peak traffic. Furthermore, 08:00 am corresponds to average/normal expected wind power. Therefore the study in this chapter, for example, Figure 6.12, was designed to show the achievable data rata in different part of city at a given hour that had normal vehicular traffic and normal wind power. Conversely the study shows that at high traffic the opportunities for sleep cycles are diminished, Figure 6.12 (a). Figure 6.11 show the trend of sleep duration and data rate based on the traffic demand at each NS at hour 08:00 a.m. It reiterates that with increase in traffic demand the sleep duration decreases. On the contrary, in the case of rate adaptation, the data rate decreases with decrease in traffic demand. Figure 6.12, shows the effect of distributing extra RE energy on sleep duration and

data rate at different NSs in the city. Here, the hourly available renewable energy improves the average piece delay if there is any excess energy available after meeting the minimum energy requirement. The MILP Model 6.3 distributes the excess RE such that the busy NSs would receive higher excess RE to improve the users experience in terms of improved download (service) rates. Models 6.4 allocate the extra available renewable (wind) energy to each NS according to their IoT Packets demand to provide best achievable delay. NSs with higher traffic demand would have more RE energy at their disposal and respectively low sleep durations.

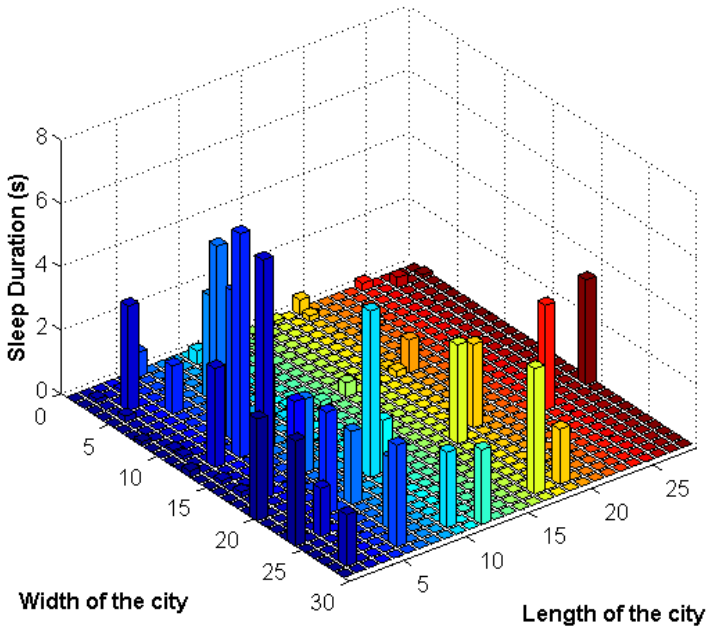


(a)

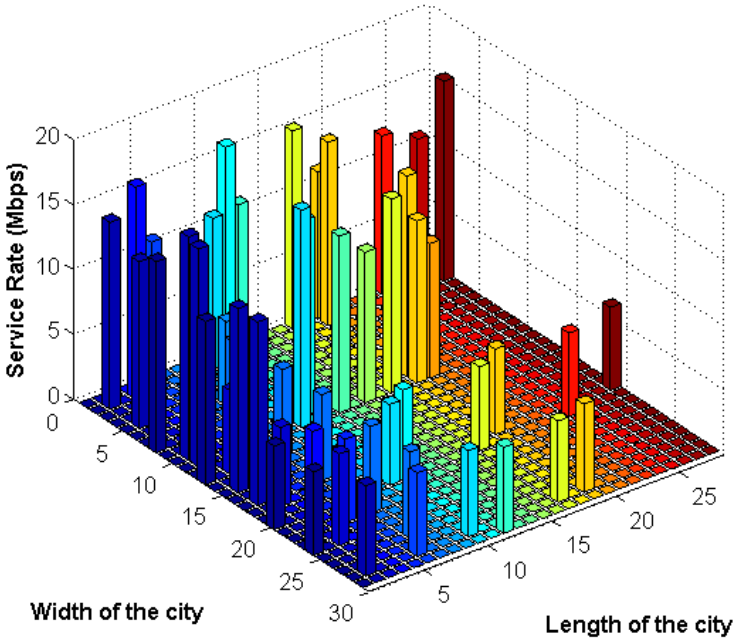


(b)

Figure 6.11 : (a) Sleep duration (b) data rate at each NS for a selected hour (Hour 08:00).



(a)



(b)

Figure 6.12: (a) Sleep duration (b) data rate at each NS (in different part of the city) for a selected hour (08:00 A.M).

As mentioned earlier, the maximum energy consumption corresponds to the consumption of installed NSs without rate-adaptation. The non-rate adaptive (traditional) NSs operate at maximum data rate giving rise to lowest piece delay. Similarly, minimum energy consumption refers to the NRE consumption of the rate adaptive NSs operating at minimum data rates. Such data rates are bounded by the maximum average piece delay requirement for respective traffic loads. The optimal energy consumption refers to the situation where the MILP Model 4 selects the busy NSs with the lowest data rates and tries to maximise the data rates with the available excess RE. This effectively minimises the carbon footprint and also yields maximum improvement in the achieved data rate. For the NSs with random sleep cycles, maximum energy is consumed when no sleep is operated and minimum energy is consumed when sleep cycles are operated with maximum duration. Similarly, optimal energy consumption refers to the load dependent renewable energy distribution, where the NSs operate with variable sleep duration to reduce the waiting delay. Thus, the average piece delay significantly reduces during most of the day (Figure 6.13).

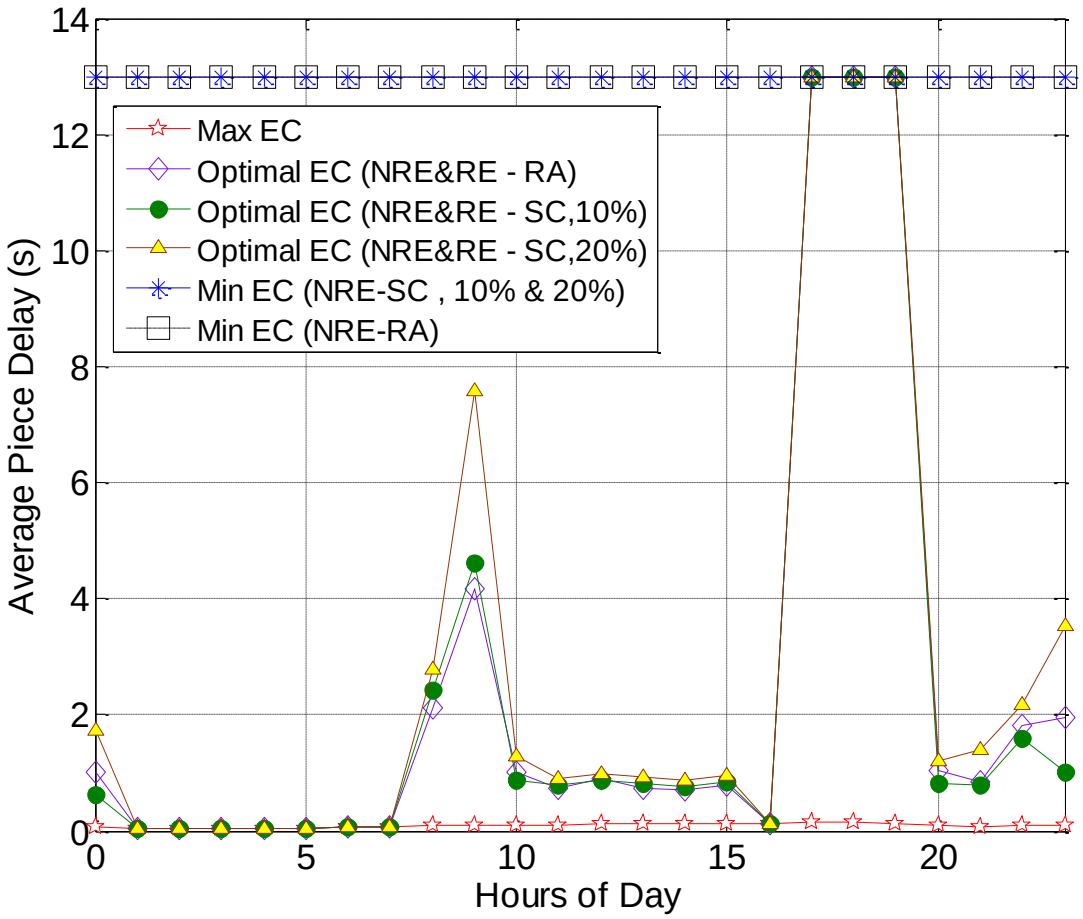


Figure 6.13 : Average piece delay at different location of city for varying hours of the day.

6.8 Conclusions

In this chapter, we proposed sleep enabled and rate adaptive Nano Servers for energy efficient delivery of IoT packets in a smart city vehicular environment. A rate adaptive NS operates at variable data rate according to the available transmission energy while a sleep enabled NS is switched to sleep mode to save transmission energy. At the first stage, the proposed MILP Model 1

optimised the number of (traditional non-rate adaptive) NSs for a typical smart city scenario by a continuous selection process of NSs (transient optimisation) according to varying hourly traffic. The number and location optimisation resulted in 49% energy savings compared to a non-optimised setup. A further 74% energy saving was obtained through transient optimisation in which, the MILP model drives the switching of non-active NSs selectively based upon spatial-temporal variation of the traffic.

Model 6.1 corresponds to transient optimisation which saved up to 81% and 80% energy respectively, for a wake up overhead energy of 10% and 20% of the maximum transmission energy. Similarly Model 6.2 incorporated rate adaptive NSs, which saved up to 78% of the transmission power without exceeding the average piece delay limit. The above savings in non-renewable energy was complemented by renewable energy available from a wind farm. The challenge of using renewable energy lied in its judicious distribution to the rate adaptive NSs such that the overall QoS can be maximised. Given this aim, we proposed MILP Models 6.3 and 6.4 which first replaced the non-renewable energy with the available renewable energy. The optimisation models then distributed the excess renewable energy to improve QoS. The load dependent distribution of the excess renewable energy improved the individual data rate of the NSs in a way that most of the piece downloads would experience minimum average piece delay. The MILP Model 6.3 improved the average piece delay by 82% and minimised the carbon footprint by 96% in the IoT packet delivery in

our smart city vehicular environment. Such a rate adaptive NSs would be essential for future Internet of Things. The MILP Model 6.4 reduced the average piece delay by 81% and 83% for the corresponding wake up overhead energy of 10% and 20% of the maximum transmission energy, and minimised the carbon footprint by 98% in the same scenario. All the MILP models are validated by genetic algorithms, which are very efficient and produced results comparable to MILP. For the case of location optimisation the deviation is 4%, while for the case of optimum usage of non-renewable energy and renewable energy mix, the deviation is about 6% (for both rate adaptation and sleep cycles).

Finally, it is to be noted that both sleep cycles and rate adaptation are effective with smaller devices like nano servers in Fog computing, where the switching overhead is acceptable unlike the case of cloud architecture. Rate adaptation is further desirable as it does not incur switching overhead and thereby maintains continuous service.

Chapter 7: Energy Efficient Nano-Servers with request dropping probability

7.1 Introduction

In this chapter, we introduce energy adaptive NSs (ADP-NSs) for IoT Information Piece Delivery (IoTIPD). The energy adaptive NSs operate at variable transmission (and networking circuitry) power and therefore can operate at variable data rates. This results in variable request dropping probability (RDP) throughout the day. There are two key goals in this chapter, to determine the amount of power that can be served if the service level agreement (SLA) allows up to a given percentage of requests to be rejected, not accepted, for example 95% SLA.

The second goal is to determine the reduction in request dropping that can be achieved if extra power is available. Therefore the chapter includes three studies.

(i) MILP optimisation of number and locations of nano servers under non adaptive hardware with the goal of minimising power

consumption while accepting a certain percentage of request dropping.

(ii) Given a number of NSs (numbers and locations) in (i), if a new manufacturer is able to provide rate adaptive NS, the goal is to use the existing NS locations hired in the city, use the new rate adaptive hardware and determine the power saving by accepting the maximum acceptable request dropping.

(iii) In the third study, if additional power (e.g. renewable) is available, the goal is to determine the reduction in request dropping that can be achieved.

7.2 Nano servers for IoT Information Piece Delivery (IoTIPD)

We developed MILP models for optimising the locations of non-adaptive Nano Servers (NSs) with sufficient non-renewable energy. The request dropping probability (RDP) is thus zero in this case. We then reduce the number of NSs by re-optimising the number of Nano servers with the constraint of Request Dropping Probability (RDP), which is maintained under 5% level. This also minimises the overall energy consumption of the NSs. Note that the NSs here are non-Energy Adaptive (non-adaptive in simpler term), which means that these operate at full energy regardless of the IoT piece demand. The

formulation sets, parameters and variables for all the MILP models are defined in Table 7.1.

Table 7.1: List of Symbols.

Set	**Description**
TP	Set of traffic points
CS	Set of candidate sites
NS	Set of installed Fog computing Nano servers
$N[j]$	Set of neighbouring TPs for $NS\ n$
$N[n]$	Set of neighbouring NSs for $TPs\ j$
T	Set of time points within one hour (600 s each)
MILP Parameter	**Description**
$Bmax$	Maximum capacity of a NS (27 Mbps)
RDP	Maximum acceptable request dropping probability (i.e. 0.05)
$Ndmax$	Maximum number of simultaneous downloads from a NS (9)
$drmax_{nt}$	Maximum data rate at $NS\ s$ at time t (i.e. 27 Mbps)
$Pmax_{nt}$	Maximum power consumption of at $NS\ s$ at time t (30 W [98])
$Pidle_{nt}$	Operational power consumption of $NS\ s$at time t 23 W [119]
$Ptxmax_{nt}$	Maximum transmission power consumption of a NS at time t given by $Pmax_{nt} - Pidle_{nt} = 7\ W$ [118]
Pw	Hourly available wind power

P_{nt}	Adaptive power consumption by $NS\ n$ at time t
R_{jt}	Request data rate at $TP\ j$ at time t
A	A constant, set to 600
σ	A constant, set to 10
MILP Variables	**Description**
Ptx_{nt}	Adaptive transmission power of $NS\ n$ at time t
$Ptxmin_{nt}$	Minimum transmission power of $NS\ n$ at time t according maximum acceptable delay
α_n	Equals 1 if $NS\ n$ is on, equals 0 otherwise
R_{njt}	Request data rate between $NS\ n$ and TP j at time t
R_n	Total Request data rate at $NS\ n$ at time t
δ_{njt}	Equals 1 if $NS\ n$ is transmitting IoT content to $TP\ j$, equals 0 otherwise
Pre_{nt}	The part of the available renewable power used by $NS\ n$ at time t
dr_{nt}	Adaptive data rate at $NS\ n$ at time t
Index	**Description**
n	Index of Nano Servers (NS)
j	Index of Traffic Point (TP)
t	Index of Time Point (T)

The NSs in IoTD receive information piece requests from the vehicles in range (D). The MILP model receives these inputs. Thus, the traffic demand (Mbps) varies at each of these NSs due to vehicular mobility. The effect of discretisation is reduced by adopting a large number of TPs. In 7.4, we proposed heuristic which accounts for the vehicular mobility. The corresponding performance results were found to be not sensitive to the described discretisation approach, which made the MILP modelling possible. The IoT Information Piece demand at each TP matrices are used as inputs to the proposed MILP model to find the optimum locations of the NSs. The objective is to minimise the total power consumption of the NSs over a given time, which yields minimum number of NSs required.

7.2.1 Non-adaptive Nano Servers powered by non-renewable grid energy

The aim in this case is to minimise the total power consumption of the NSs over the entire time period t. The corresponding MILP model ensures that at each time point, the total traffic is served. We used the results of location optimisation Chapter 5 based on MILP model in Chapter 4.

7.2.2 Non-adaptive Nano Servers with request dropping probability powered by non-renewable grid energy

In this case, we want to optimise the location and number of nano-servers to minimise the total power consumption by considering the request dropping probability.

The objective of the MILP model is

$$\boldsymbol{Minimize}$$

$$\sum_{t\epsilon T}\sum_{n\epsilon NS}\alpha_n(Ptxmax_{nt} + Pidle_{nt}) \tag{7.1}$$

$$\forall\, n \epsilon NS, \forall\, t\epsilon T$$

Subject to the following constraints:

$$\frac{\sum_{n\epsilon NS}(R_{nt} - Bmax)}{\sum_{j\epsilon TP} R_{jt}} \leq RDP \tag{7.2}$$

$$\forall\, n \epsilon NS, \forall\, j\epsilon TP, , \forall\, t\epsilon T$$

where

$$\sum_{n\epsilon NS} R_{nt} = \sum_{j\epsilon TP} \lambda_{jt} \tag{7.3}$$

$$\forall\, n \epsilon NS, \forall\, j\epsilon TP, \forall\, t\epsilon T$$

$$R_{jt} = \sum_{n\epsilon NS} R_{njt} \tag{7.4}$$

$$\forall\, n \epsilon NS, \forall\, j\epsilon TP, \forall\, t\epsilon T$$

$$\frac{R_{njt}}{drmax_{nt}} \leq \delta_{njt} A \tag{7.5}$$

$$\forall\, n \epsilon NS, \forall\, j\epsilon TP, \forall\, t\epsilon T$$

$$\frac{R_{njt}}{drmax_{nt}} \geq \delta_{njt} \tag{7.6}$$

$$\forall\, n \epsilon NS, \forall\, j\epsilon TP, \forall\, t\epsilon T$$

$$\sum_{j \epsilon TP} \delta_{njt} \geq \alpha_n \quad (7.7)$$

$$\forall\, n \,\epsilon\, NS, \forall\, j \epsilon TP, \forall\, t \epsilon T$$

$$\sum_{j \epsilon TP} \delta_{njt} \leq \alpha_n A \quad (7.8)$$

$$\forall\, n \,\epsilon\, NS, \forall\, j \epsilon TP, \forall\, t \epsilon T$$

The request dropping probability constraint is given by Equation (7.2). Equation (7.2) ensures that the traffic in the city does not exceed the collective capacity of nano-servers by more than 5% and equation (7.3) ensures that all the requests in the city are allocated to the nano-servers where the total amount of requests is R_{jt} and the variable R_{njt} defines the requests dealt by the $NS\ n$ of the $TP\ j$ at time t. Equation (7.4) ensures that at each time point the capacity of a NS is not violated. If a NS in an TP has already reached its full capacity, the model proposes installation of another nano-server to serve the remaining requests. Therefore, the download rate is defined as Equation (7.5). Equation (7.5) ensures that if the number of requests is non-zero between $NS\ n$ and j, $\delta_{njt} = 1$. Equation (7.6) ensures that if the requests is zero between NS n and $TP\ j$ i.e. $R_{njt} = 0$, then there is no connection. Hence, $\delta_{njt} = 0$. Equations (7.7) and (7.8) ensure that if there is a connection between $NS\ n$ and $TP\ j$ at a time point t, then $NS\ n$ is switched ON ($\alpha_n = 1$) otherwise set $\alpha_n = 0$, switch Off the $NS\ n$.

7.2.3 Energy Adaptive Nano Server powered by non-renewable grid energy with a service level agreement on Request Dropping Probability

In this scenario, we introduce energy adaptive nano-servers where transmission energy is adaptive. The same number and location of nano-servers are used with hardware upgrade to allow us to study the impact of rate adaptation because new full optimisation may yield different number and locations of adaptive NS. We introduced a simplistic (linear) relationship between transmission energy and the capacity of a NS, the NS can operate at lower capacity to reduce power consumption and serve the total requests at the same time. This occurs only when the demand is low. However, it is achieved at the expense of higher request dropping probability (RDP). Therefore, the RDP constraint needs to ensure that the dropping is kept equal or under 5%. As the main objective of this model is to minimise energy consumption, the RDP is pushed to the maximum (5%) to maximise energy saving.

The objective is to maximise the requests allocated to a NS n and is given as:

$$\boldsymbol{Maximise}$$

$$\frac{\sum_{n\epsilon NS}(R_{nt} - dr_{nt})}{\sum_{j\epsilon TP} R_{jt}} \tag{7.9}$$

$$\forall\, n \,\epsilon\, NS, \forall\, j\epsilon TP, \forall\, t\epsilon T$$

This MILP includes constraints (7.2) and (7.3) along with the constraints below

$$R_{nt} - Bmax \geq -A(1 - \beta_{nt}) \quad (7.10)$$

$$\forall\, n \in NS, \forall\, t \epsilon T$$

$$R_{nt} - Bmax \leq A\, \beta_{nt} \quad (7.11)$$

$$\forall\, n \in NS, \forall\, t \epsilon T$$

$$dr_{nt} = (1 - \beta_{nt}) R_{nt} + \beta_{nt} Bmax \quad (7.12)$$

$$\forall\, n \in NS, \forall\, t \epsilon T$$

$$Ptx_{nt} = \frac{dr_{nt}}{Bmax} Ptxmax_{nt} \quad (7.13)$$

$$\forall\, n \in NS, \forall\, t \epsilon T$$

$$P_{nt} = Ptx_{nt} + Pidle_{nt} \quad (7.14)$$

$$\forall\, n \in NS\,, \forall\, t \epsilon T$$

$$\sum_{t \epsilon T} \sum_{n \epsilon NS} Ptx_{nt} + Pidle_{nt} \quad (7.15)$$

$$\forall\, n \in NS\,, \forall\, t \epsilon T$$

Equations (7.10) and (7.10) set the binary variable $\beta_{nt} = 1$ if the load R_{nt} at NS n exceeds its capacity ($RDP > 0$) where the NS needs to remove the part of R_{nt} which is more than $Bmax$ and serve the rest, else $\beta_{nt} = 0$.

The transmission power needs to be calculated by including a constraint that all traffic has to be served. Equation (7.12) and (7.13) ensure each $NS\ s$, adapts its data rate to serve the total traffic and calculates the transmission power needed. Therefore, the total power consumption of $NS\ n$ at time t is given as Equation (7.14) and the total power consumption of NSs in the city is given as Equation (7.15).

7.2.4 Energy Adaptive Nano Server with Optimum Request dropping probability powered by non-renewable and renewable grid energy

In this scenario, the goal is to minimise the RDP, by injecting more power into the network. We utilise renewable (wind) grid energy for this purpose to reduce carbon footprint. According to the availability of wind energy [94], the MILP model developed adjusts the capacity of NSs to its appropriate value to serve IoT piece requests with lower energy consumption. The main constraints defined in this scenario correspond to (i) replacing non-renewable energy (NRE) with RE (ii) improving RDP by using extra RE, in the case of excess RE.

The objective is:

$$\boldsymbol{Minimise} \tag{7.16}$$

$$\frac{\sum_{n\epsilon NS}(R_{nt} - dr_{nt})}{\sum_{j\epsilon TP} R_{jt}}$$

$$\forall\, n \,\epsilon\, NS, \forall\, j\epsilon TP, \forall\, t\epsilon T$$

Subject to:

$$dr_{nt} = (1 - \beta_{nt}) dr_{nt} + \beta_{nt} Bmax \tag{7.17}$$

$$\forall\, n \,\epsilon\, NS, \forall\, t\epsilon T$$

$$dr_{it} - dr_{kt} \geq \sigma\,(R_{it} - R_{kt}) \tag{7.18}$$

$$\forall i \;\&\; k \,\epsilon\, NS, i \neq k, \forall\, t\epsilon T$$

$$Ptx_{nt} = \frac{dr_{nt}}{Bmax_{nt}} \; Ptxmax_{nt} \tag{7.19}$$

$$\forall \; n \; \epsilon \; NS, \forall \; t \epsilon T$$

$$Pre_{nt} \leq Ptx_{nt} + Pidle_{nt} \tag{7.20}$$

$$\forall \; n \; \epsilon \; NS \; , \forall \; t \epsilon T$$

$$\sum_{n \; \epsilon \; NS} Pre_{nt} \leq Pw \tag{7.21}$$

$$\forall \; n \; \epsilon \; NS \; , \forall \; t \epsilon T$$

This MILP include equations (7.2), (7.3), (7.10), (7.11) and (7.14) as constraints and equation (7.12) is modified as Equation (7.17). Equation (7.18) is used to allocate a higher data rate to a NS that has higher traffic request to serve. Then the transmission power consumption is calculated as Equation (7.19) and renewable usage constraints are given as Equation (7.20).

7.3 HEURISTICS

We develop heuristics to validate the 3 different scenarios involving in this chapter Nano Servers (NSs). The corresponding algorithms are shown in Flowcharts (7.1) to (7.3): IoT Information Piece Delivery. The first case computes the request dropping probability (RDP) for the optimised number of nano servers with given traffic (IoT Information Piece) demand and maintains RDP below 5%, as described in Figure 7.1. In the second case, we reduce the energy consumption by keeping the RDP at the 5% level. Unlike the previous cases, we consider energy adaptive NSs in this case (Figure7.2). Finally, in the

third case we introduce renewable power along with non-renewable power and optimise the RDP (Figure 7.3). We used the results of developed custom heuristic in Chapter 5 for comparison.

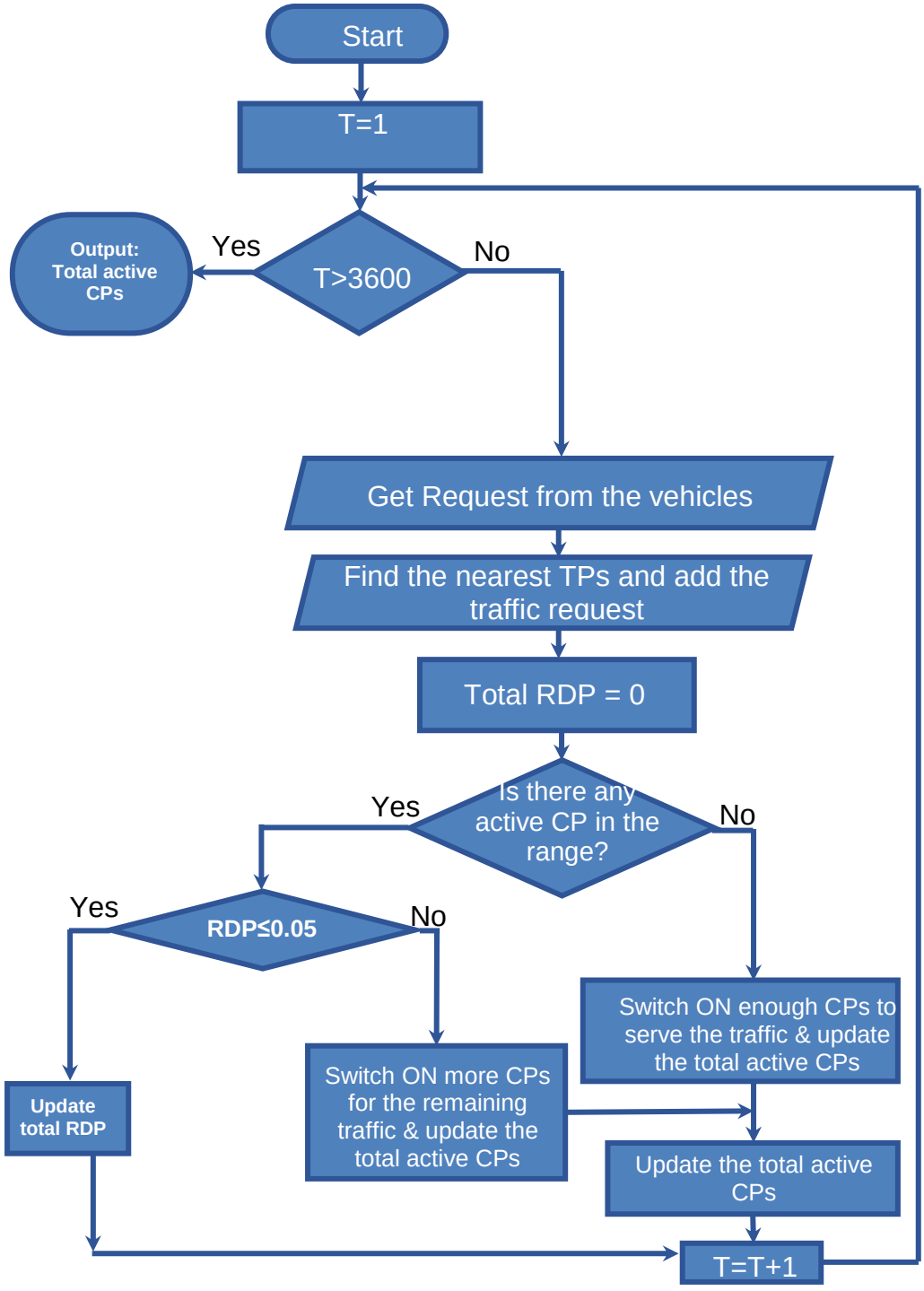


Figure 7.1: Location Optimisation with RDP.

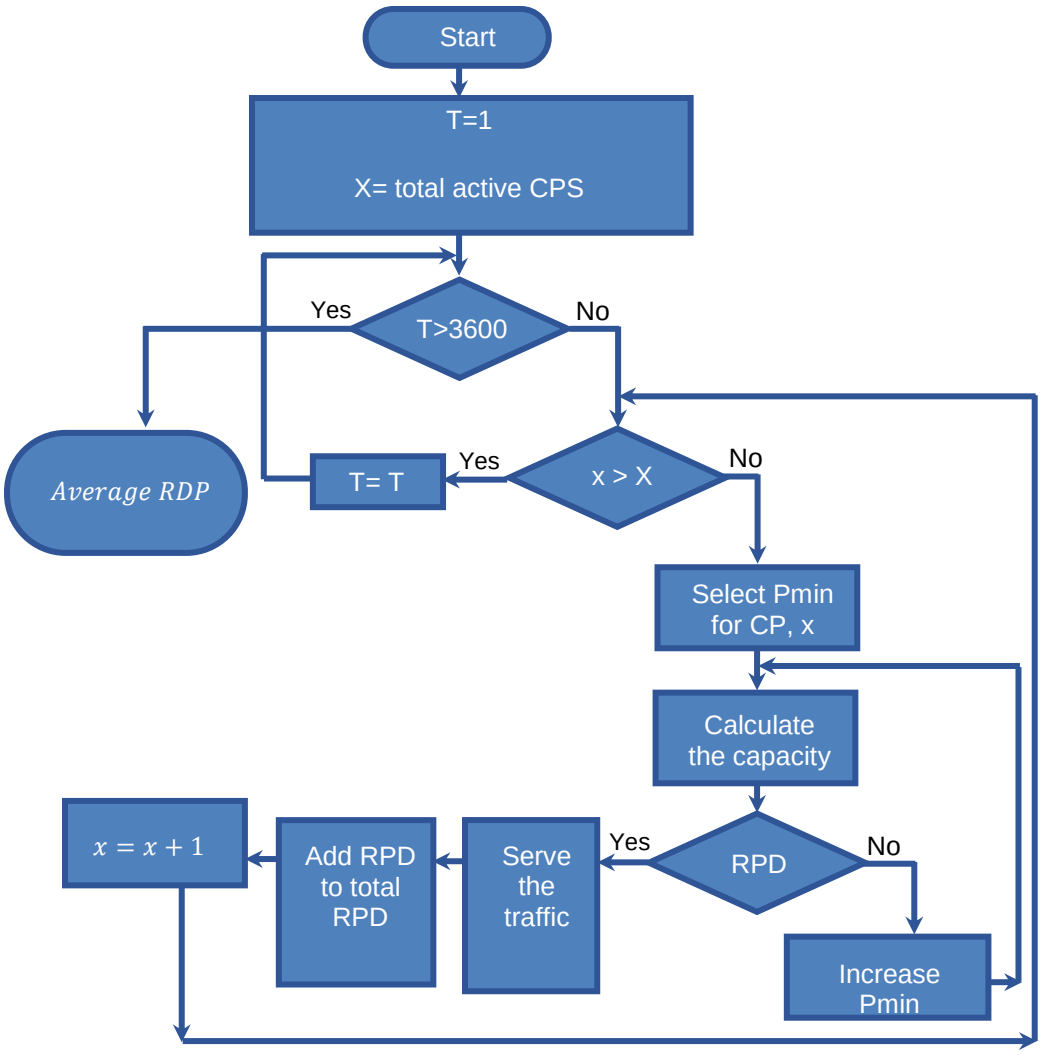


Figure 7.2: Location Optimisation with MAX RDP.

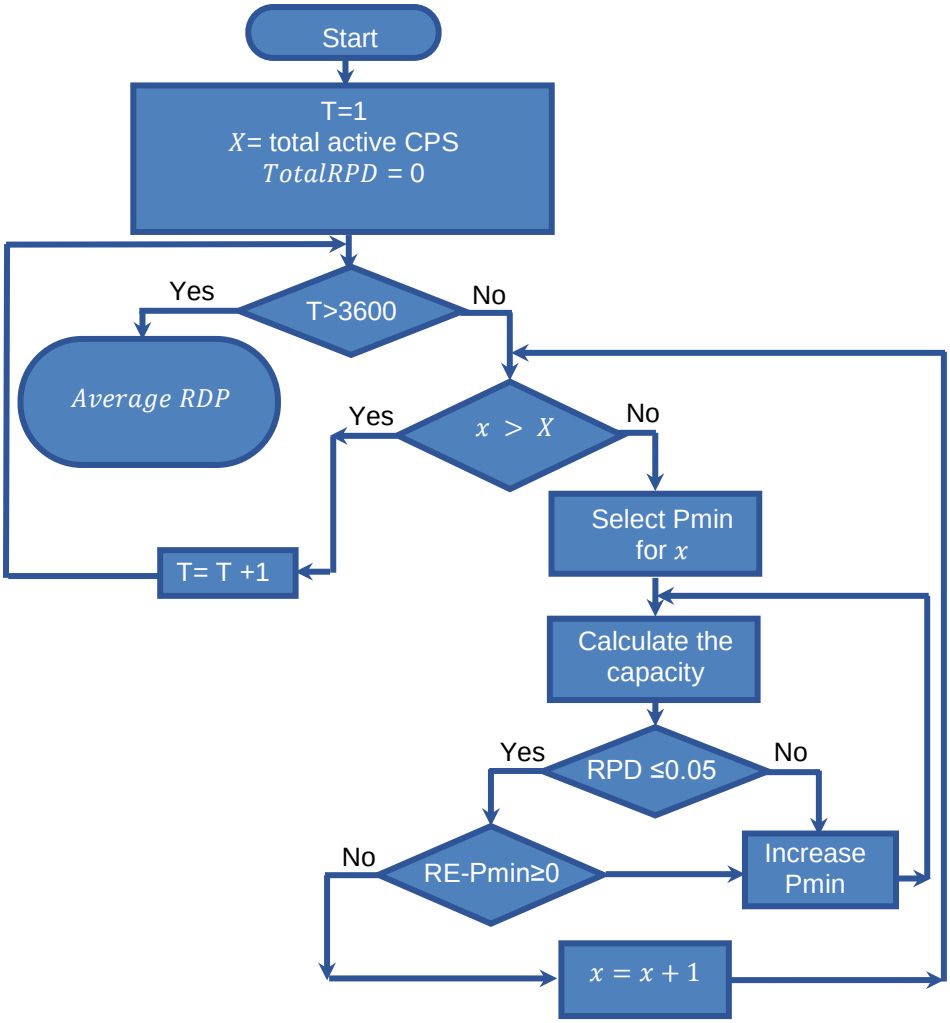


Figure 7.3: Location Optimisation with RE and OPT RDP.

7.5 PERFORMANCE EVALUATION

In this section, we discuss the performance evaluation of the IoT nano Servers in an information piece delivery scenario to the city (vehicular) users. We consider five different cases with different combination of non-renewable and renewable energy for non-adaptive and adaptive nano servers. The first case is for illustration purpose, which considers fixed non-renewable grid energy available to the nano servers throughout the day. In this case the locations are non-optimised. The effect of variation of piece demand on request dropping probability (RDP) is shown in Figure 7.4. The piece demand for varying hours of the day is derived from the city vehicular traffic profile. Thus, when the demand is low, all the demands are served. At the peak hours, we observe 5% dropping as the nano servers are incapable of serving all the requests.

Figure 7.5 shows the hourly variation of the number of optimised nano servers. As the number of NSs is reduced, the RDP increases accordingly. Our proposed MILP model ensures that the peak RDP does not exceed 5%. Whereas, the heuristic algorithm operates with the instantaneous knowledge available, this does not restrict installing additional nano servers. Thus, a relatively higher number of installed NSs is observed in the case of the heuristic algorithm. Overall, the results of the heuristics follow the same trends as those of the MILP optimisation model.

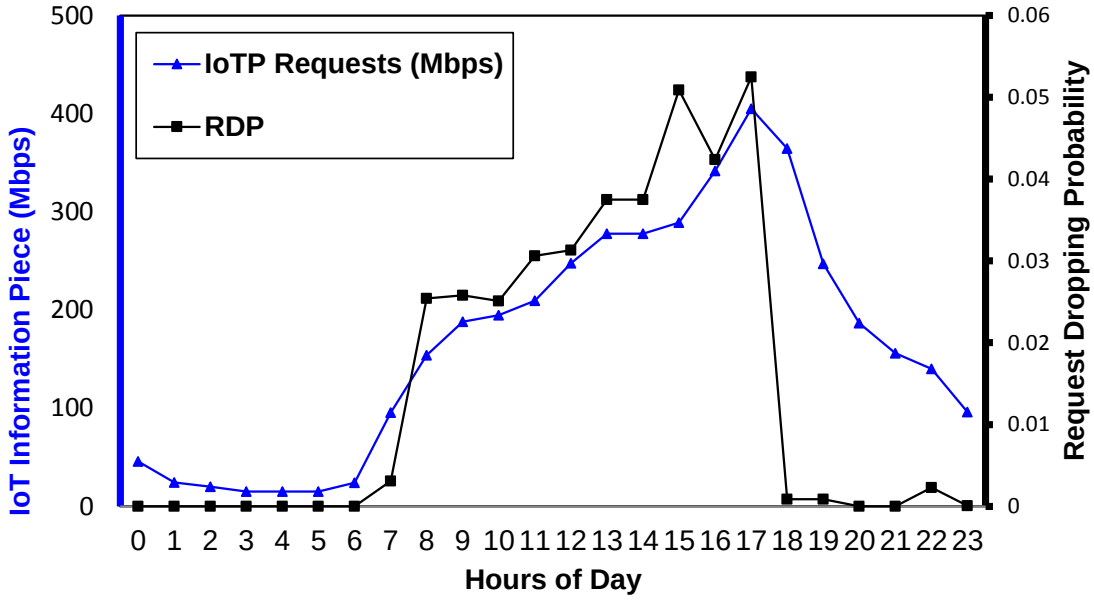


Figure 7.4: IoT information piece demand and Request dropping probability in smart city scenario.

In Figure 7.5, the Blue curve (lines for MILP and dots Heuristics) shows the variation of the number of NSs for varying hours of the day without any RDP. The RED curves (dashed for MILP and circled for Heuristic) shows the variation of RDP for varying hours of the day, which does not cross the peak value of 5%. Evidently, the number of NSs is more in the former case. This occurs due to the non-adaptive NSs in presence of non-renewable energy. All the curves predominantly follow the hourly variation of piece demand.

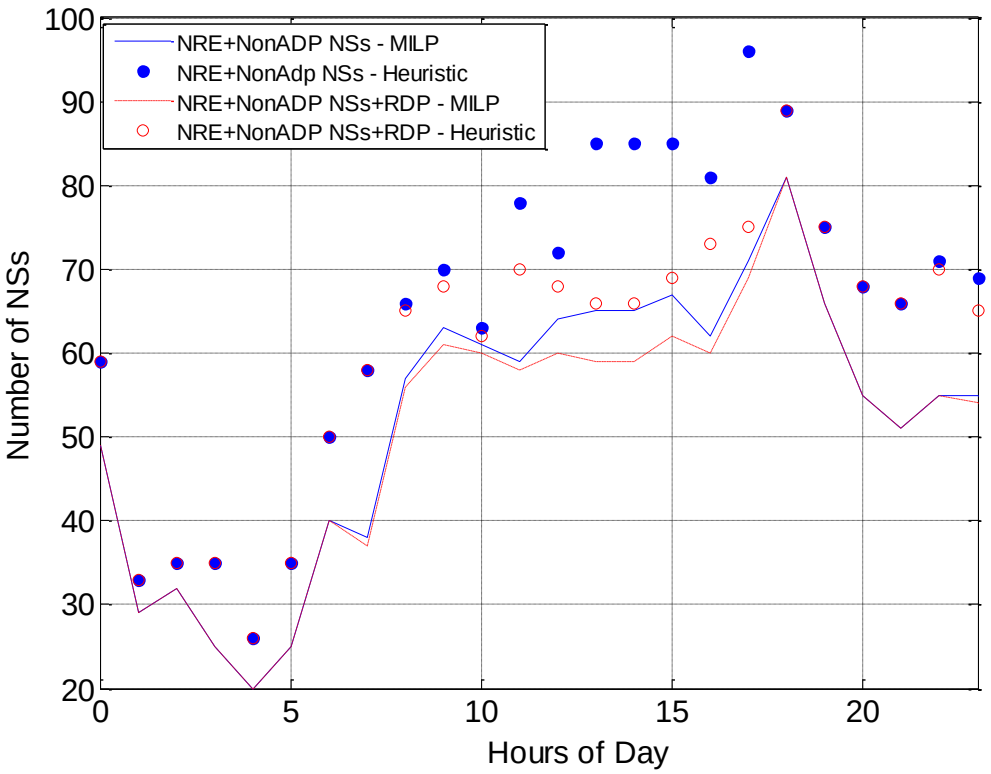


Figure 7.5: Number of Nano servers for varying hours of the day.

Figure 7.6 shows different cases corresponding to different combinations of non-adaptive, adaptive NSs with Renewable and Non-Renewable energy. The bold lines are MILP results and the circles are Heuristic results. The RED curve shows the RDP with non-adaptive NSs. Therefore, it follows piece demand variation. A minor decrement in dropping (at 19-21 hours) is experienced with adaptive NSs and Non-Renewable energy (BLUE curves). However, with the availability of renewable energy (GREEN curves), which is also varying in nature, the RDP is reduced considerably throughout the day. This also ensures minimum use of NRE. In the case of adaptive NSs (Blue curves), we minimise NRE, which pushes the RDP to the maximum (5%). When the RE is sufficient

to turn on the adaptive NSs, they operate at the lowest rate. The RDP, therefore, is the highest (5%). When the RE is sufficient, the RDP reduces to zero. When the RE available to an adaptive NSs varies between its minimum and maximum operating level, the RDP varies between zero and 5%. Here, we use the minimum NRE at each hour of the day. So, the RDP is optimised in that sense. Further, it is observed that the results obtained with MILP model and heuristic algorithm are in good agreement, even though the results of the heuristics does not exactly follow the trends of MILP models as the number of NSs in case of heuristics are higher than that of MILP, which has a direct effect on RDP.

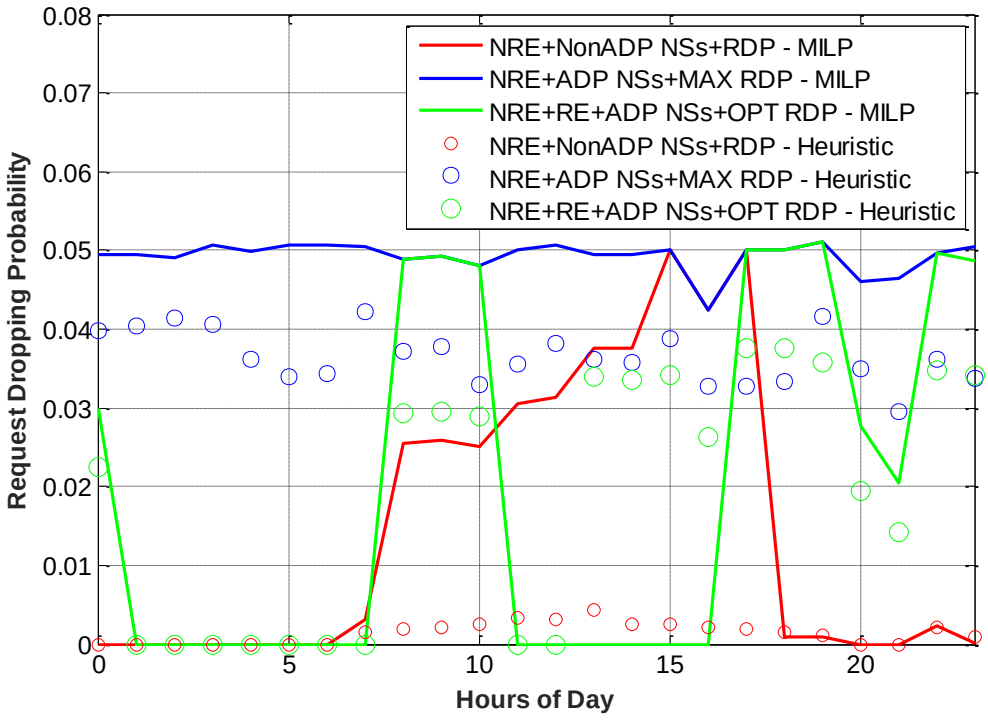
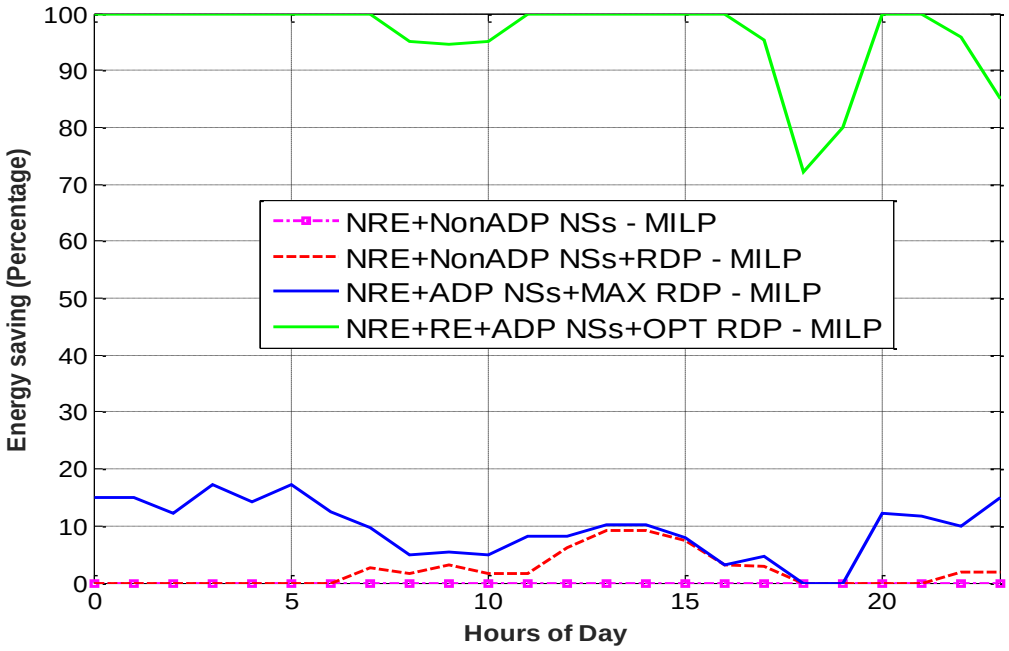
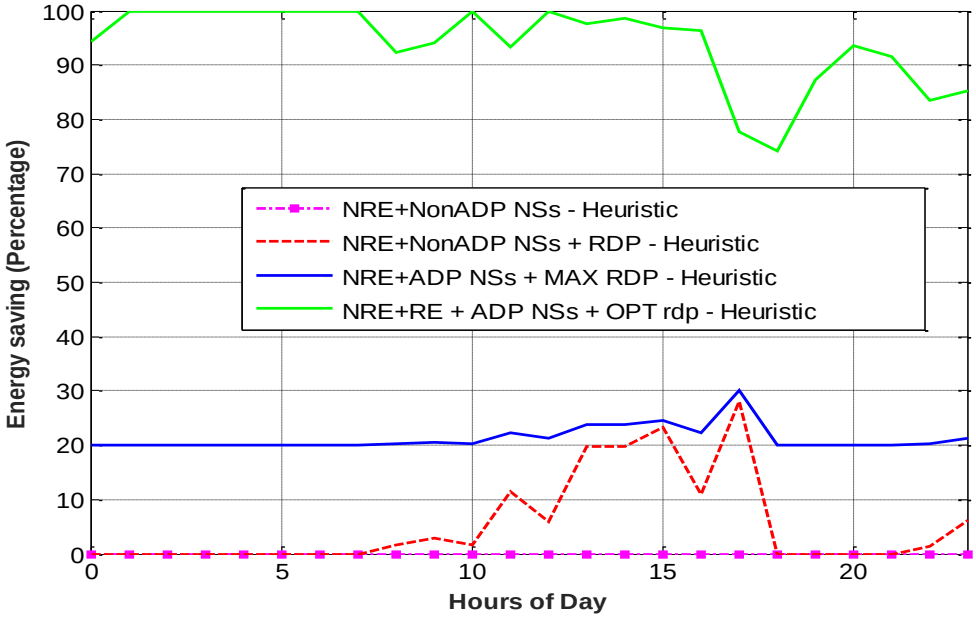


Figure 7.6: Request Dropping Probability (RDP) for varying hours of the day.

Figure 7.7 illustrate the total energy savings obtained by the respective models at each hour of the day for MILP (a) and heuristic (b). The Pink line shows the non-renewable energy savings with non-adaptive NSs (which is negligible). The red dotted line shows non-renewable energy savings with RDP, which is marginally better. The blue line shows non-renewable energy savings with maximum dropping. This is the best that can be achieved with NRE. The green line shows the best achievable energy savings with optimum dropping. This is possible with the introduction of renewable grid energy and minimisation of non-renewable energy.



(a)



(b)

Figure 7.7: Hourly variation of energy savings a) MILP b) Heuristic.

7.6 CONCLUSIONS

In this chapter, we proposed an Internet of Things Information Piece Delivery through nano servers in the city vehicular scenario. We proposed adaptive nano servers, which can operate with a variable amount of available renewable and non-renewable energy. We observed that with renewable energy, we reasonably maintained the dropping within the limit. Further, we conclude that the adaptive nano servers are much more flexible than the non-adaptive Nano servers since they maintain reasonable dropping even with insufficient energy. This is possible because the adaptive nano servers operate at variable capacity and consume variable energy. This is achieved at the expense of piece dropping. However, our study showed that dropping can be kept under an acceptable level if we judiciously use the available renewable energy.

Chapter 8: Conclusions and Summary

8.1 Conclusions and Key Findings

In this chapter, the dissertation is concluded by drawing some key findings on performance analysis and modelling of vehicular content delivery networks and their roadmap towards the domain of Internet-of-Things (IoT).

In the case of load adaptive cache points (CPs), the data rate adapts to the load, which is modelled through statistical multiplexing of a number of servers in queuing theory. It is basically a form of dynamic serving of the connections thereby operating at an adaptive data rate. This enables a CP to operate sleep cycles and save transmission energy and also maintain the average piece delay within acceptable limits. The performance result depicts 84% savings in transmission energy during off-peak hours and 33% during the whole day while fulfilling content demand in a city vehicular environment. Moreover, secondary CPs, equipped with sleep mechanism due to their lower load, have achieved higher energy savings during their operating span compared to the primary CPs. Furthermore, it was concluded that the mean sleep cycle duration has insignificant effect on the energy savings of both CPs due to the large piece size (i.e. 6 MB), long service time and low wake-up overhead.

In the next piece of work, energy adaptive CPs for vehicular content delivery network are proposed. Here, the data rate varies according to the varying transmission energy. Thus, the cache points are classified as energy adaptive instead of load adaptive. The average piece delays at the proposed CPs were dependent on the available renewable energy. For non-adaptive CPs, the whole traffic demands can only be served at the expense of higher number of active CPs. It means that if we want to keep the same number of CPs, the whole traffic demand will not be served at a number of hours. This increases the overall average piece delay at earlier hour of the day when the available wind energy is insufficient. In contrast, the non-renewable energy consumption increases with the scarcity of renewable energy while the fixed number of CPs remains active. The advantage of such CPs is in terms of acceptable average piece delay and substantial savings of non-renewable energy. The study in this chapter revealed that the proposed CPs can fully eliminate the need for non-renewable grid energy during the whole day while fulfilling content demand in a typical city vehicular environment.

In the third work, the regime of fog computing in a smart city vehicular environment is introduced. It comes under the umbrella of IoT. We considered nano servers for fog computing. As a first step, we minimised the total power consumption of the nano servers in smart city vehicular environment by optimising the number and locations of ordinary (non-adaptive) nano-servers. This is a static optimisation. Secondly, a heuristic was developed to verify the results of the MILP model. Finally, the performance of the scenario was

evaluated in terms of the number of installed and active nano-servers. The optimisation resulted in 49% energy saving compared to the non-optimised setup. A further 74% of energy savings was obtained through transient optimisation, in which the MILP model switches off the non-active nano-servers selectively based upon the spatiol-temporal variation of the traffic. This is followed by energy saving through nano servers implementing sleep cycles, which saved up to 81% and 80% transmission energy, respectively corresponding to wake up overhead energy of 10% and 20% of the maximum transmission energy. Rate adaptive nano servers saved up to 78% of transmission energy without exceeding the average piece delay limit. In the next work we introduced renewable energy driven nano servers. The challenge of using renewable energy lied in its judicious distribution to the rate adaptive nano-servers such that overall QoS can be maximised. Given this goal, we proposed two successive MILP models, which first replaced the non-renewable energy with available renewable energy and then distributed the excess renewable energy to improve QoS. Renewable energy powered rate-adaptive nano servers reduced the average piece delay by 82% and minimised the carbon footprint by 96%. Renewable energy powered, sleep cycle enabled nano server improved the average piece delay by 81% and 83% for the corresponding wake-up overhead energy of 10% and 20% of the maximum transmission energy, and minimised the carbon footprint by 98% in the same scenario.

Finally, it is to be noted that both sleep cycles and rate adaptation are effective with smaller devices like cache points, fog computing servers or nano-servers, where the switching overhead is acceptable unlike the case of cloud architecture. Rate adaptation is further desirable as it does not incur switching overhead and thereby maintains continuous service. The author believes that such techniques will become popular in the case of 5G and beyond which is presumed to include IoT and its standardisation process.

8.2 Future Work

This thesis included a number of novel contributions. Analytical studies and simulations were carried out and used to characterise these new systems and compared them to existing, established designs. The pieces of work carried out in this thesis open a number of new avenues for future research. The need for a green energy power supply has spurred an interest in energy harvesting. Harvesting solar energy, and using that as an energy supply with the techniques studied in this thesis, is an interesting research topic. The performance modelling of safety related applications together with infotainment delivery involves priority based servicing. This gives rise to the avenues of tackling the optimisation problem in a more specialised way. Further, with smart city applications various real time information need to be processed and delivered in a timely manner as requested by the users. An example of such

applications includes complete assistance on official procedures in timely and non-hazardous manner. This might be tackled using combinatorial optimisation.

References

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Communications Surveys & Tutorials*, vol. 17, pp. 2347-2376, 2015.

[2] G. Suciu, A. Vulpe, S. Halunga, O. Fratu, G. Todoran, and V. Suciu, "Smart Cities Built on Resilient Cloud Computing and Secure Internet of Things," in *2013 19th International Conference on Control Systems and Computer Science*, 2013, pp. 513-518.

[3] T. L. Willke, P. Tientrakool, and N. F. Maxemchuk, "A survey of inter-vehicle communication protocols and their applications," *IEEE Communications Surveys & Tutorials*, vol. 11, pp. 3-20, 2009.

[4] R. Popescu-Zeletin, I. Radusch, and M. A. Rigani, Vehicular-2-X communication: state-of-the-art and research in mobile vehicular ad hoc networks: Springer Science & Business Media, 2010.

[5] X. Sheng, N. F. Maxemchuk, and L. Cai, "Special Section on Vehicular Communication Networks," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 3241-3243, 2007.

[6] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, et al., "Vehicular Networking: A Survey and Tutorial on Requirements, Architectures, Challenges, Standards and Solutions," *IEEE Communications Surveys & Tutorials*, vol. 13, pp. 584-616, 2011.

[7] J. Guo and N. Balon, "Vehicular ad hoc networks and dedicated short-range communication," *unpublished*, September, vol. 22, 2006.

[8] C. Han, M. Dianati, R. Tafazolli, R. Kernchen, and X. Shen, "Analytical Study of the IEEE 802.11p MAC Sublayer in Vehicular Networks," *IEEE Transactions on Intelligent Transportation Systems,* vol. 13, pp. 873-886, 2012.

[9] H. Kawashima, "Japanese perspective of driver information systems," *Transportation,* vol. 17, pp. 263-284, 1990.

[10] M. Tubaishat, P. Zhuang, Q. Qi, and Y. Shang, "Wireless sensor networks in intelligent transportation systems," *Wireless communications and mobile computing,* vol. 9, pp. 287-302, 2009.

[11] T. Aizono, M. Shioya, and K. Sano, "Wireless method and system for providing navigation information," ed: Google Patents, 2003.

[12] M. T. Qadri and M. Asif, "Automatic number plate recognition system for vehicle identification using optical character recognition," in *2009*

International Conference on Education Technology and Computer, pp. 335-338, 2009.

[13] W. Chen, R. K. Guha, T. J. Kwon, J. Lee, and Y. Y. Hsu, "A survey and challenges in routing and data dissemination in vehicular ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 11, pp. 787-795, 2009.

[14] R. Stanica, E. Chaput, and A.-L. Beylot, "Properties of the MAC layer in safety vehicular Ad Hoc networks," Communications Magazine, IEEE, vol. 50, pp. 192-200, 2012.

[15] Y. Khaled, B. Ducourthial, and M. Shawky, "IEEE 802.11 performances for inter-vehicle communication networks," in Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st, pp. 2925-2929, 2005.

[16] Hao, W. *Analysis and design of vehicular networks*. Diss. Ph. D. thesis, Georgia Institute of Technology, Atlanta, USA, 2005.

[17] M. L. Sichitiu and M. Kihl, "Inter-vehicle communication systems: a survey," IEEE Communications Surveys & Tutorials, vol. 10, pp. 88-105, 2008.

[18] C. Harsch, A. Festag, and P. Papadimitratos, "Secure position-based routing for VANETs," in *2007 IEEE 66th Vehicular Technology Conference*, pp. 26-30, 2007.

[19] S. k. Gupta and S. Khara, "Technique to improve the file transfer outcomes between road side unit and vehicles in vehicular ad-hoc networks," in *International Conference on Computing, Communication & Automation*, pp. 360-366, 2015.

[20] S. Yousefi, M. S. Mousavi, and M. Fathy, "Vehicular Ad Hoc Networks (VANETs): Challenges and Perspectives," in *2006 6th International Conference on ITS Telecommunications*, pp. 761-766, 2006.

[21] Y. Wang, G. Attebury and B. Ramamurthy, "A survey of security issues in wireless sensor networks," in IEEE Communications Surveys & Tutorials, vol. 8, no. 2, pp. 2-23, Second Quarter 2006.

[22] D. Jiang and L. Delgrossi, "IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, pp. 2036-2040, 2008.

[23] J. Jakubiak and Y. Koucheryavy, "State of the art and research challenges for VANETs," in Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE, pp. 912-916, 2008.

[24] A. Prakash and R. Tripathi, "Vehicular ad hoc networks toward Intelligent Transport Systems," in *TENCON 2008 - 2008 IEEE Region 10 Conference*, pp. 1-6, 2008.

[25] T. D. Nguyen, O. Berder, and O. Sentieys, "Energy-Efficient Cooperative Techniques for Infrastructure-to-Vehicle Communications," *IEEE Transactions on Intelligent Transportation Systems,* vol. 12, pp. 659-668, 2011.

[26] H. Zhang, Y. Ma, D. Yuan, and H. H. Chen, "Quality-of-Service Driven Power and Sub-Carrier Allocation Policy for Vehicular Communication Networks," *IEEE Journal on Selected Areas in Communications,* vol. 29, pp. 197-206, 2011.

[27] P. Kolios, V. Friderikos, and K. Papadaki, "Ultra Low Energy Store-Carry and Forward Relaying within the Cell," in *2009 IEEE 70th Vehicular Technology Conference Fall*, pp. 1-5, 2009.

[28] S. H. Wu, C. M. Chen, and M. S. Chen, "An Asymmetric and Asynchronous Energy Conservation Protocol for Vehicular Networks," *IEEE Transactions on Mobile Computing,* vol. 9, pp. 98-111, 2010.

[29] W. Kumar, A. Muhtar, B. R. Qazi, and J. M. H. Elmirghani, "Energy and QoS Evaluation for a V2R Network," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, 2011, pp. 1-5.

[30] W. Kumar, S. Bhattacharya, B. R. Qazi, and J. M. H. Elmirghani, "An energy efficient double cluster head routing scheme for motorway vehicular networks," in *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 141-146.

[31] L. Wischhof, “Self-Organizing Communication in Vehicular Ad Hoc Networks,” Diss. Ph. D. thesis, TU Hamburg, 2007.

[32] L. Wischhof, A. Ebner, and H. Rohling, "Information dissemination in self-organizing intervehicle networks," *IEEE Transactions on Intelligent Transportation Systems,* vol. 6, pp. 90-101, 2005.

[33] M. Torrent-Moreno, M. Killat, and H. Hartenstein, "The challenges of robust inter-vehicle communications," in *VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference*, pp. 319-323, 2005.

[34] J. J. Blum, A. Eskandarian, and L. J. Hoffman, "Challenges of intervehicle ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems,* vol. 5, pp. 347-351, 2004.

[35] K. Matheus, R. Morich, and A. Lübke, "Economic background of car-to-car communication," in Proceedings of the 2. Braunschweiger Symposium Informationssysteme für mobile Anwendungen. Braunschweig, 2004.

[36] B. R. Qazi, H. Alshaer, and J. Elmirghani, "Development of a Motorway Simulator for Vehicular Multimedia Communications," in Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th, 2008, pp. 1-5.

[36] B. R. Qazi, H. Alshaer, and J. M. H. Elmirghani, "Development of a Motorway Simulator for Vehicular Multimedia Communications," in *2008 IEEE 68th Vehicular Technology Conference*, pp. 1-5, 2008.

[37] M. Iavadi, S. Habib, and M. Hannan, "Survey on Inter-Vehicle Communication Applications: Current Trends and Challenges," *Information Technology Journal*, vol. 12, pp. 243-250, 2013.

[38] Y. Toor, P. Muhlethaler, A. Laouiti, and A. D. L. Fortelle, "Vehicle Ad Hoc networks: applications and related technical issues," *IEEE Communications Surveys & Tutorials,* vol. 10, pp. 74-88, 2008.

[39] H. Menouar, F. Filali, and M. Lenardi, "A survey and qualitative analysis of MAC protocols for vehicular ad hoc networks," *IEEE Wireless Communications*, vol. 13, pp. 30-35, 2006.

[40] A. Sugiura and C. Dermawan, "In traffic jam IVC-RVC system for ITS using Bluetooth," *IEEE Transactions on Intelligent Transportation Systems,* vol. 6, pp. 302-313, 2005.

[41] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *IEEE Communications Magazine*, vol. 44, pp. 74-82, 2006.

[42] K. Bilstrup, “A survey regarding wireless communication standards intended for a high-speed vehicle environment,” Technical Report IDE 0712, Halmstad University, Sweden, 2007.

[43] M. Torrent-Moreno, Inter-vehicle communications: Achieving safety in a distributed wireless environment: Challenges, systems and protocols: Universitätsverlag Karlsruhe, 2007.

[44] M. Li, Z. Yang, and W. Lou, "CodeOn: Cooperative Popular Content Distribution for Vehicular Networks using Symbol Level Network Coding," *IEEE Journal on Selected Areas in Communications,* vol. 29, pp. 223-235, 2011.

[45] A. Capone, "Energy and mobility: Scalable solutions for the mobile data explosion”, *Green Touch June Open Forum*. 2012.

[46] J. Oueis, E. C. Strinati, S. Sardellitti, and S. Barbarossa, "Small Cell Clustering for Efficient Distributed Fog Computing: A Multi-User Case," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, 2015, pp. 1-5.

[47] R. Hussain, J. Son, H. Eun, S. Kim, and H. Oh, "Rethinking Vehicular Communications: Merging VANET with cloud computing," in *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings*, pp. 606-609, 2012.

[48] S. Manish Kumar and A. Kaur, "A survey on Vehicular Cloud Computing and its security," in *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, pp. 67-71, 2015.

[49] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," pp. 13-16. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing,* pp. 13-16, 2012.

[50] I. Stojmenovic, "Fog computing: A cloud to the ground support for smart things and machine-to-machine networks," in *2014 Australasian Telecommunication Networks and Applications Conference (ATNAC)*, pp. 117-122, 2014.

[51] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, et al., "Millimeter wave mobile communications for 5G cellular: It will work!,", *IEEE Access*, vol. 1, pp. 335-349, 2013.

[52] S. Sen, J. Rexford, and D. Towsley, "Proxy prefix caching for multimedia streams," in *INFOCOM '99. Eighteenth Annual Joint Conference of the*

IEEE Computer and Communications Societies. Proceedings. IEEE, pp. 1310-1319 vol.3, 1999.

[53] M. Johnson, L. De Nardis, and K. Ramchandran, "Collaborative content distribution for vehicular ad hoc networks," in *Allerton conference on communication, control, and computing*, pp. 751-760, 2006.

[54] D. Zhang and C. K. Yeo, "Enabling Efficient WiFi-Based Vehicular Content Distribution," *IEEE Transactions on Parallel and Distributed Systems,* vol. 24, pp. 479-492, 2013.

[55] U. Shevade, Y.-C. Chen, L. Qiu, Y. Zhang, V. Chandar, M. K. Han, et al., "Enabling high-bandwidth vehicular content distribution," in *Proceedings of the 6th International COnference*, 2010, p. 23, 2010.

[56] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, p. 10 pp. vol.2, 2000.

[57] C. Peng, C. Chen-Nee, and L. Xin, "Energy-aware node placement in wireless sensor networks," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, pp. 3210-3214 Vol.5, 2004.

[58] E. Callaway, "Low power consumption features of the IEEE 802.15. 4/ZigBee IR-WPAN standard," *Mini-tutorial, ACM Sensys,* vol. 3, pp. 5-7, 2003.

[59] Y. T. Hou, S. Yi, H. D. Sherali, and S. F. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Transactions on Wireless Communications,* vol. 4, pp. 2579-2590, 2005.

[60] N. Li, Y. Xu, and S.l. Xie, "A power-saving protocol for ad hoc networks," in *Proceedings. 2005 International Conference on Wireless Communications, Networking and Mobile Computing,* pp. 808-811, 2005.

[61] H. Chih-Shun and T. Yu-Chee, "Cluster-based semi-asynchronous power-saving protocols for multi-hop ad hoc networks," in *IEEE International Conference on Communications,* pp. 3166-3170 Vol. 5, 2005.

[62] W. Kumar, S. Bhattacharya, B. R. Qazi, and J. M. H. Elmirghani, "A Vacation-Based Performance Analysis of an Energy-Efficient Motorway Vehicular Communication System," *IEEE Transactions on Vehicular* Technology, vol. 63, pp. 1827-1842, 2014.

[63] G. Benelli, R. Fantacci, G. Giambene, and C. Ortolani, "Voice and data transmissions with a PRMA-like protocol in high propagation delay

cellular systems," IEEE Transactions on Vehicular Technology, vol. 49, pp. 2126-2147, 2000.

[64] Guo, Song, and Oliver Yang. "Minimum energy multicast routing for wireless ad-hoc networks with adaptive antennas." Network Protocols, *ICNP 2004. Proceedings of the 12th IEEE International Conference on. IEEE, 2004.*

[65] W. Fisher, M. Suchara, and J. Rexford, "Greening backbone networks: reducing energy consumption by shutting off cables in bundled links," in *Proceedings of the first ACM SIGCOMM workshop on Green networking*, pp. 29-34, 2010.

[66] M. Gupta and S. Singh, "Greening of the Internet," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 19-26, 2003.

[67] G. Fettweis and E. Zimmermann, "ICT energy consumption-trends and challenges," in *Proceedings of the 11th International Symposium on Wireless Personal Multimedia Communications*, Vol. 2. No. 4. p. 6, 2008.

[68] H. Idjmayyel, W. Kumar, B. R. Qazi, and J. M. H. Elmirghani, "Saving energy with QoS for vehicular communication," in *2011 Eighth International Conference on Wireless and Optical Communications Networks*, pp. 1-5, 2011.

[69] H. Idjmayyel, B. R. Qazi, and J. M. H. Elmirghani, "Position Based Routing Protocol for a City Environment," in *2010 Fourth International Conference on Next Generation Mobile Applications, Services and Technologies*, pp. 174-179, 2010.

[70] H. Idjmayyel, W. Kumar, B. R. Qazi, and J. M. H. Elmirghani, "Energy and QoS - A new perspective in a city vehicular communication network," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pp. 327-331, 2011.

[71] W. Kumar, S. Bhattacharya, B. R. Qazi, and J. M. H. Elmirghani, "A Vacation-Based Performance Analysis of an Energy-Efficient Motorway Vehicular Communication System," *IEEE Transactions on Vehicular Technology,* vol. 63, pp. 1827-1842, 2014.

[72] M. Rajaratnam and F. Takawira, "Two moment analysis of channel reservation in cellular networks," in *1999 IEEE Africon. 5th Africon Conference in Africa (Cat. No.99CH36342)*, pp. 257-262 vol.1, 1999.

[73] Transport Planning. Report. City of Saskatoon. [Online]. Available: http://www.saskatoon.ca/DEPARTMENTS/Infrastructure%20Services/Transportation/TransportationPlanning/Documents/2009%20Traffic%20Characteristic%20Report.pdf (Visited December 2012)

[74] B. R. Qazi and J. M. H. Elmirghani, "M-PRMA protocol for vehicular multimedia communication," *International Journal of Communication Networks and Distributed Systems,* vol. 4, pp. 28-48, 2009.

[75] A. Capone. Energy and mobility: Scalable solutions for the mobile data explosion. GreenTouch. Politecnico di Milano. [Online]. Available: http://www.greentouch.org/uploads/documents/4%20Antonio%20Capone_SEW%20Capone%20presentation.pdf (Visited April 2014)

[76] S. E. Elayoubi, L. Saker, and T. Chahed, "Optimal control for base station sleep mode in energy efficient radio access networks," in *2011 Proceedings IEEE INFOCOM*, pp. 106-110, 2011.

[77] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks," in *2009 IEEE 70th Vehicular Technology Conference Fall*, pp. 1-5, 2009.

[78] A. U. Joshi and P. Kulkarni, "Vehicular WiFi access and rate adaptation," In *ACM SIGCOMM Computer Communication Review*. vol. 40. no. 4., pp. 423-424, 2010.

[79] J. Camp and E. Knightly, "Modulation rate adaptation in urban and vehicular environments: cross-layer implementation and experimental evaluation," Networking, IEEE/ACM Transactions on, vol. 18, pp. 1949-1962, 2010.

[80] P. Shankar, T. Nadeem, J. Rosca, and L. Iftode, "CARS: Context-Aware Rate Selection for vehicular networks," in *2008 IEEE International Conference on Network Protocols*, pp. 1-12, 2008.

[81] D. Zhibin, Z. Zenghua, J. Quan, Z. Lianfang, S. Yantai, and O. Yang, "Energy-efficient rate adaptation for outdoor long distance WiFi links," in *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 271-276, 2011.

[82] H. Idjmayyel, "Energy Efficient Vehicular Networks in a City Environment," Diss. Ph. D. thesis, University of Leeds, 2013.

[83] BBC iPlayer [Online]. Available: http://www.bbc.co.uk/iplayer/(Visited April 2014)

[84] C. V. N. Index. Global Mobile Data Traffic Forecast Update, 2013–2018 [Online].Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html (Visited August 2015).

[85] F. Thouin and M. Coates, "Video-on-Demand Networks: Design Approaches and Future Challenges," *IEEE Network,* vol. 21, pp. 42-48, 2007.

[86] N. I. Osman, T. El-Gorashi, and J. M. Elmirghani, "Caching in green IP over WDM networks," *Journal of High Speed Networks*, vol. 19, pp. 33-53, 2013.

[87] W. Xiaolei, B. Yanan, L. Xin, and N. Zhisheng, "On the design of relay caching in cellular networks for energy efficiency," in *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2011, pp. 259-264, 2011.

[88] Cisco Visual Networking Index: Forecast and Methodology, 2009–2014. [Online]. Available: http://large.stanford.edu/courses/2010/ph240/abdul-kafi1/docs/white_paper_c11-481360.pdf (Visited July 2013).

[89] R. Hodges and W. White, "Go green in ICT," *Green Tech News,* 2008.

[90] B. R. Qazi, "Performance evaluation of vehicular networks," Diss. Ph. D. thesis, University of Leeds, 2010.

[91] P. Higgins and A. M. Foley, "Review of offshore wind power development in the United Kingdom," in Environment and Electrical Engineering (EEEIC), 12th International Conference on, 2013, pp. 589-593.

[92] Muskulus, Michael, and Jakob Mann. "Selected Papers from the European Wind Energy Association 2014, Barcelona." IET Renewable Power Generation 9.8, pp. 865-866, 2015.

[93] Lupo, Sara, and Aristides E. Kiprakis. "The impact of renewable energy resources on the electricity prices of the United Kingdom." European Energy Market (EEM), 13th International Conference on the. IEEE, 2016.

[94] Burgess, Paul. "Variation in light intensity at different latitudes and seasons, effects of cloud cover, and the amounts of direct and diffused light." Forres, UK: Continuous Cover Forestry Group. Available online at http://www. ccfg. org. uk/conferences/downloads/P_Burgess. pdf. 2009.

[95] Belfkira, Rachid, Lu Zhang, and Georges Barakat. "Optimal sizing study of hybrid wind/PV/diesel power generation unit." Solar Energy 85.1 pp. 100-110, 2011.

[96] Department for Environment Food & Rural Affairs, UK-AIR: Air Information Resource. [Online]. Available: http://uk-air.defra.gov.uk/. (Visited February 2015).

[97] J. Bird, “Higher engineering mathematics,” Routledge, 2007.

[98] Propeller Wind Turbine. [Online]. Available: ftp://ftp.pasco.com/Support/Documents/english/ET/ET-8776/Propeller%20Wind%20Turbine%20Mnl%20(ET-8776).pdf(Visited February 2015)

[99] Department for Transport, Leeds traffic profile. [Online]. Available: http://www.dft.gov.uk/ (Visited February 2013).

[100] Aruba products [Online]. Available: WWW.ARUBANETWORKS.COM. (Visited January 2015).

[101] I. Haratcherev, M. Fiorito, and C. Balageas, "Low-Power Sleep Mode and Out-Of-Band Wake-Up for Indoor Access Points," in *2009 IEEE Globecom Workshops*, pp. 1-6, 2009.

[102] The Smart City Market Opportunities for the UK [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/249423/bis-13-1217-smart-city-market-opportunties-uk.pdf (Visited January 2015)

[103] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," IEEE Internet of Things Journal, vol. 1, pp. 22-32, 2014.

[104] P. Bellavista, G. Cardone, A. Corradi, and L. Foschini, "Convergence of MANET and WSN in IoT urban scenarios," *IEEE Sensors Journal*, vol. 13, pp. 3558-3567, 2013.

[105] C. Perera, C. H. Liu, S. Jayawardena, and M. Chen, "A survey on internet of things from industrial market perspective," IEEE Access, vol. 2, pp. 1660-1679, 2014.

[106] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," *IEEE Communications Surveys & Tutorials*, vol. 16, pp. 414-454, 2014.

[107] F. J. Martinez, J. C. Cano, C. T. Calafate, and P. Manzoni, "CityMob: A Mobility Model Pattern Generator for VANETs," in *ICC Workshops - 2008 IEEE International Conference on Communications Workshops*, pp. 370-374, 2008.

[108] A. Kansal, F. Zhao, J. Liu, N. Kothari, and A. A. Bhattacharya, "Virtual machine power metering and provisioning," in *Proceedings of the 1st ACM symposium on Cloud computing*, pp. 39-50, 2010.

[109] F. Sheikh, H. Fazal, F. Taqvi, and J. Shamsi, "Power-aware server selection in nano data center," in *2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*, pp. 776-782, 2015.

[110] D. T. and P. Viswanath, "*Fundamentals of Wireless Communication*," Cambridge University Press, 2005.

[111] K. TrafficCom. Smarter vehicles, safer roads MCNU R1551. [Online]. Available: http://ww1.prweb.com/prfiles/2008/10/22/915544/MCNUR1551.PDF. (Visited May 2013)

[112] AMPL FOR RESEARCH [Online]. Available: http://ampl.com/BOOKLETS/amplcplex110userguide.pdf. (Visited November 2013)

[113] B. R. Johnson, "New numerical methods applied to solving the one-dimensional eigenvalue problem," *The Journal of Chemical Physics*, vol. 67, pp. 4086-4093, 1977.

[114] A. Eiger, K. Sikorski, and F. Stenger, "A bisection method for systems of nonlinear equations," *ACM Transactions on Mathematical Software (TOMS)*, vol. 10, pp. 367-377, 1984.

[115] S. Boyd, V. Balakrishnan, and P. Kabamba, "A bisection method for computing the H∞ norm of a transfer matrix and related problems," *Mathematics of Control, Signals and Systems*, vol. 2, pp. 207-219, 1989.

[116] S.-y. GE, J.-f. LI, and B.-g. ZHANG, "Optimal Configuration of Sectional Switches in Distribution Network Based on Bisection Method [J]," *Power System Technology*, vol. 13, p. 013, 2007.

[117] H. S. Wilf, "A global bisection algorithm for computing the zeros of polynomials in the complex plane," *Journal of the ACM (JACM)*, vol. 25, pp. 415-420, 1978.

[118] K. Shafique and M. Shah, "A noniterative greedy algorithm for multiframe point correspondence," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, pp. 51-65, 2005.

[119] L. M. Schmitt, "Theory of genetic algorithms," *Theoretical Computer Science*, vol. 259, pp. 1-61, 2001.

[120] W. Kumar, S. Bhattacharya, B. Qazi, and J. M. Elmirghani, "A Vacation-based Performance Analysis of an Energy-Efficient Motorway Vehicular Communication System," *IEEE Transactions on Vehicular Technology*., vol. 63, pp. 1827-1842, May 2014.

[121] I. Haratcherev, M. Fiorito, and C. Balageas, "Low-Power Sleep Mode and Out-Of-Band Wake-Up for Indoor Access Points," in 2009 *IEEE GLOBECOM Workshops*, pp. 1-6, 2009.

[122] S. K. Bose, “An Introduction to Queuing Systems,” Kluwer Academic /Plenum Publishers, 2002.

Appendices

Appendix: Algorithms

In this Appendix, we describe our proposed algorithms that are used to validate the models and are used for simulation to analyse and evaluate the corresponding work in the thesis.

Algorithm 1: Load Adaptive CP

Input: $Lambda$, $MeanSleepDuration$, $MaxConnections(c) \in CP$, $n \in Pieces$, $PieceSize, VarianceOfConnections(c)$

Output: $Delay,\ NumberOfSleepSleepDuration, ServiceDuration$

for all *time = 1, 2, 3, …., Tsim* **do**

1 ***initialize*** $IAT[piece\ (n)] = exprand(MeanInterArrivalTime)$;

2 ***Compute***

$de - Neg(MeanOfConnections(c), VarianceOfConnections(c), PieceSize)$

3 **if** $IAT[piece\ (n)] \leq 0$ **then**

4 generate Request[piece (n)];

record time [Request[piece (n)]];

5 add Request[piece (n)] to CP(c). Queue;

6 **if** *CP(c).sleep = 1* **then**

7 $SleepDuration\ [CP(c)] - -$;

if $SleepDuration\ [CP(c)] \leq 0$ **then**

8 $CP(c).sleep = 0$;

9 $CP(c).busy = 0$;

else if $CP(c)$ is $Busy$ **then**

10 $ServiceDuration - -$;

if ServiceDuration ≤ 0 **then**

11 $CP(c).busy = 0$;

12 $CP(c).sleep = 0$;

Algorithm 1: Load Adaptive CP

13 **else if** $CP(c).Queue > 0$ **then**

initialize $ServiceDuration = exprand(det - Neg)$;

14

start $Download\ [piece[(n)]$;

15

record $time[Download\ [piece(n)]$;

16

remove Request[piece (n)] from CP(c). Queue;

$CP(c).busy = 1;0$

17

$CP(c).sleep = 0;$

18

Compute Delay[piece (n)];

else

19

Set

20

$SleepDuration[cp] = exprand\ (MeanSleepDuration);$

$NumberOfSleep(c) + +;0$

21

$CP(c).sleep = 1;$

22

$CP(c).busy = 0;$

end if

23

end if

24

else

$IAT[piece\ (n)] - -;$

25

end if

26

$time + +;$

end if

27

end for

Algorithm 2: Generate Traffic

Input: $\forall\ v \in Vehicle$, $\forall j \in TP$, $c \in CS$, $simulation\ time$

Output: $Traffic[TP[j], t]$

```
for all t = 1,2,.., simulation time do
    for all j ∈ TP do
        for all f ∈ FS do
            create neighbourhood list of TP[j];
        end for
    end for
    for all v ∈ Vehicle do
        Set waiting time = exprand (mean inter arrival time (τ))
        if waiting time ≤ 0 then
            generate Request[IoT Content];
            PieceNumber = 1000
            record time[Request[IoT Content]];
        end if
        if PieceNumber > 0 then
            find nearest TP[j] for v;
            Traffic[TP[j]] + +;
            record time[Request[piece n]];
            PieceNumber—
        end if
    end for
    waiting time − −
end for
```

Algorithm 3: CDN Heuristic

Input: $Lambda$,$MaxNoOfConnections$, $AvailableWindEnergy$, $RequestedEnergy$

Output: $Number\ of\ CPs$

1 **for all** $t = 1, 2, 3, \ldots T_{Sim}$ **do**

2 **for all** TPs **do**

3 $Calculate\ \ NeededNoOfConnections$

4 **NRE without Non Adaptive CPs CASE:**

5 **if** any CP in Neighbouring of TP is ***Active & Available Capacity>0***

6 **if** $NeededNoOfConnections \leq$ ***Available Capacity*** **then**

7 Serve the traffic demand

8 **else**

9 ***Turn On*** enough CPs form the neighboring

10 and Served the traffic demand

11 end if

12 **else**

13

14 ***Turn On*** enough CPs form the neighboring

15 and Served the traffic demand

16 **end if**

17 **RE without Non Adaptive CPs CASE:**

18 **if** any CP in Neighbouring of TP is ***Active & Available Capacity>0***

19 **if** $NeededNoOfConnections \leq$ ***Available Capacity*** **then**

20 $Serve\ traffic\ demand$

21 **else if** $RequestedEnergy(c) \leq AvailableWindEnergy(c)$ **then**

22 ***Turn On*** enough CP and Served the traffic demand

23 **end if**

24 **else if** $RequestedEnergy(c) \leq AvailableWindEnergy(c)$ **then**

25 ***Turn On*** enough CP and Served the traffic demand

26 **else**

27 $Drop\ all\ the\ remain\ requests$

28 **end if**

29 **RE with Adaptive CPs CASE:**

Algorithm 3: CDN Heuristic

```
30      if any CP in Neighbouring of TP is Active & Available Capacity>0
31          if NeededNoOfConnections ≤ Available Capacity then
32              Serve traffic demand
33          else
34              Calculate Available Capacity of each CP
35              Turn On enough CPs form the neighboring
                and Served the traffic demand
36          end if
37      else
38          Calculate Available Capacity of each CP
39          Turn On enough CPs form the neighboring
40          and Served the traffic demand
41      end if
42      Compute Delay
    end for
    t++
end for
```

Algorithm 4: A Rate Adaptive Nano Server

Input: $\forall \lambda_n \in T_\lambda$, E , $fs \in NS$

Output: $AvgDelay$, $ns.Queue$

1. **for all** $time$ = 1, 2, 3, …., T_{sim} **do**
2. Set $IATPiece(i) = Random(MeanIATPiece)$;
3. **if** $IATPiece(i) \leq 0$ **then**
4. ***Generate*** $Request[i]$;
5. ***Record*** $Time[Request[i]]$;
6. ***Add*** $Request[i]$ to $ns.Queue$;
7. **if** $ns.Busy$ = 1 **then**
8. $ServiceDuration - -$;
9. **if** $ServiceDuration \leq 0$ **then**
10. ***Compute*** $AvgDelay[i]$;
11. ***Remove*** $Request[i]$ from $ns.Queue$;
12. **if** $ns.Queue > 0$ **then**
13. ***Compute*** $dataRate$ from E;
14. ***Compute*** $ServiceDuration$;
15. **else**
16. $ns.Busy$ = 0;

Algorithm 4: A Rate Adaptive Nano Server

17. **end if**
18. **end if**
19. **end if**
20. **else**
21. $IATPiece--$;
22. **end if**
23. $time$++;
24. **end for**

Algorithm 5: Location Optimisation - Custom Heuristic

Input: $\forall v \in Vehicle, Traffic[TP[j]], \forall j \in TP, n \epsilon NS, MaxNoOfConnection, NSNeighbourOfTP[j]$

Output: $NoOfActiveNSs[j]$

1. **for all** $t \in T$ **do**
2. **for all** $j \in TP$ **do**
3. **if** $Traffic[TP[j]] > 0$
4. Calculate $NoOfNeededNSs$
5. **if** $NoOfNeededNSs[j] > 0$
6. **for all** $\boldsymbol{TP}[\boldsymbol{k}] \in TPNeighbourOfTP[j]$ when $Traffic[TP[k]] > 0$
7. **Controller :**
8. $find$ $\boldsymbol{SharedNS}$
9. $Activate$ $\boldsymbol{SharedNS}$ $if\ it's\ not\ active$
10. $use\ available\ connections\ of$ $\boldsymbol{active\ SharedNS}$
11. **if** $total\ available\ connection\ of$ $\boldsymbol{SharedNSs}$ $are\ not\ enough$
12. $\boldsymbol{active}$ $\boldsymbol{Non\_SharedNSs}$ in_range
13. **end if**
14. **end for**
15. **end if**
16. **end if**
17. **end for**

Algorithm 6: Location Optimisation -Bisection

Input: $Traffic[TP[j]], \forall j \in TP, n \epsilon NS, NSNeighbourOfTP[j]$

Output: $NoOfActiveNSs[j]$

18. **for all** $t \in T$ **do**
19. **for all** $\boldsymbol{TP}$ **do**
20. **if** $Traffic[TP[j]] > 0$
21. $\boldsymbol{Calculate\ \ NoOfNeededNSs}$
22. **if** $NoOfNeededNSs[j] > 0$
23. **for all** $\boldsymbol{TP\ do}$
24. $use\ available\ connections\ of\ \boldsymbol{active\ NS}$
25. **if** $total\ available\ connections\ of\ \boldsymbol{active\ NSs}\ are\ not\ enough$
26. $Randomly\ select\ 2\ set\ of\ NSs\ \&$
27. $\boldsymbol{Calculate}\ amount\ of\ Traffic\ served(\boldsymbol{F})$
28. $L = \boldsymbol{NSs}\ set\ with\ lowest\ \boldsymbol{F}$
29. $R = \boldsymbol{NSs}\ set\ with\ Highest\ \boldsymbol{F}$
30. **end if**
31. **end for**
32. **for all** $\boldsymbol{TP\ do}$
33. **if** $\boldsymbol{F}(R) - \boldsymbol{F}(L) > 0.01$
34. $Middle = (L - Option + R - Option)/2$
35. **if** $\boldsymbol{F}(Middle) - \boldsymbol{F}(L) > \boldsymbol{F}(R) - \boldsymbol{F}(Middle)$

Algorithm 6: Location Optimisation -Bisection

36. $L = Middle$
37. **else**
38. $R = Middle$
39. **end if**
40. **end if**
41. **end for**
42. **end if**
43. **end if**
44. **end for**
45. **end for**

Algorithm 7: Location Optimisation - Greedy

Input: $Traffic[TP[j]], \forall j \in TP, n \epsilon NS, MaxNoOfConnection,$

$NSNeighbourOfTP[j]$

Output: $NoOfActiveNSs[j]$

```
1.  for all t ∈ T do
2.      for all TP do
3.          if Traffic[TP[j]] > 0
4.              Calculate NoOfNeededNSs
5.              if NoOfNeededNSs[j] > 0
6.                  for all TP do
7.                      Randomly Select i(100) sets of NSs
8.                      A= Select half of i that serve more traffic
9.                      if all of elemts of A are served the whole traffic
10.                         Sort A & select the smallest
11.                         NoOfActiveNSs = the smallest i
12.                     else
13.                         for all i in A do:
14.                             Add one NS to set i
15.                             Check the amount of traffic that is served by i
16.                             if i is served the whole traffic
17.                                 Add i to A
```

Algorithm 7: Location Optimisation - Greedy

```
18.                                    i++
19.                                end if
20.                            end for
21.                        end if
22.                    end for
23.                end if
24.            end if
25.        end for
26.    end for
```

Algorithm 8: Location Optimisation – Genetic

Input : $Traffic[TP[j]], \forall j \in TP, n \epsilon NS,\ NSNeighbourOfTP[j]$

Output: $NoOfActiveNSs[j]$

```
1.  for all TP do
2.      if Traffic[TP[j]] > 0
3.          Generate Initial Population :
4.          for population size (g=400) do
5.              Generate first :
6.              Optimized Locations[g] = Randomly activate NSs for TP[j]
7.          end for
8.          for No Of iteration (i) do
9.              for population size do
10.                 Evaluate Fitness of Each Individual :
11.                 Delay[g]=Compute delay for TP j
12.                 check validity:
13.                 if Delay[g] =< Max Acceptable Delay
14.                     Optimized Locations [g] is valid & Fitness [g]=
15.                     Energy Saving [g]
16.                 else
17.                     Optimized Locations[g] is not valid & remove it
18.                     from generation
19.                 end if
```

Algorithm 8: Location Optimisation – Genetic

```
        end for
        Selection:
        Sort fitness[g] & choose m(g/2) number of gen with
        higher fitness
        Crossover:
        for n(g/2) do
            Choose parents X, Y from Selected generation
            Optimized Locations[n] = Crossover( X, Y )& add to
            Selected population [g]
        end for
        Mutation:
        for 10% of population size do
            Optimized Locations[g] =
            Mutate (Optimized Locations[g])
        end for
    end for
     Sort fitness[g] & select Optimized Locations with the
    highest fitness
     if division of top two highest fitness is less than 5%
        Break for loop
     end if
end if
```

Algorithm 9: Optimised usage of NRE & RE - Genetic Algorithm

Input: ∀ $ActiveNS[TP[j]], NoOfPieceRequest\ @TP[j]$, *No Of iteration (i)* $AvailableRenewableEnergy$

Output: $AveragePieceDelay$

1. **for all** $time$ **do**
2. **if** $AvailableRenewableEnergy > MinEnergyRequired$
3. **Generate Initial Population :**
4. **for** *population size (g=400)* **do**
5. Extra RE$[g]$ = $\boldsymbol{Randomly}\ distribute\ ExtraRE\ for\ all\ TPs$
6. **end for**
7. **for** *No Of iteration (i)* **do**
8. **for** *population size* **do**
9. **Evaluate Fitness of Each Individual (SC and RA):**
10. $AveragePieceDelay\ [g]$= Average of (Compute
11. delay for each $\boldsymbol{TP}$)
12. **Check validity:**
13. **if** $AveragePieceDelay\ [g]$=< *MaxAcceptableDelay*
14. $ExtraRE[g]$ is valid & Fitness[g] =
15. $AveragePieceDelay\ [g]$
16.

Algorithm 9: Optimised usage of NRE & RE - Genetic Algorithm

else

$ExtraRE[g]$ is not valid & remove it from population

end if

end for

Selection:

Sort fitness[g] & choose $m(g/2)$ number of gen with lower fitness

Crossover:

for $n(g/2)$ **do**

Choose parents X, Y from Selected population

OptimisedLocations[n] = Crossover(X, Y)& add to Selected $ExtraRE[g]$

end for

Mutation:

for 10% of *population size* **do**

OptimisedLocations[g] = Mutate ($ExtraRE[g]$)

end for

end for

Sort fitness[g] & select *ExtraRE* with the lowest fitness

end if

end for

Algorithm 10: IoT Information Piece Delivery with RDP

Input: $IoT\ Information\ Piece\ (IoTP)\ Requests,\ AvailableWindEnergy,\ NearbyNSs$

Output: $Number\ of\ NSs, EnergySaving, PDP$

23 **for all** $t_1 = 1, 2, 3, \ldots T_{Simulation}$ **do**

24 **for all** NSs **do**

25 **NRE with Non Energy Adaptive NSs CASE:**

26 **if** $IoTP\ Requests > 0$ **then**

27

28 ***Find*** $NearbyNSs$

29 **if** $Any\ NearbyNSs\ is\ ON \& AvailableCapacity > 0$

30 **if** $IoTP\ Requests \leq Available\ Capacity$

31 $Serve\ the\ IoTP\ Requests$

32 **else**

33 ***Turn ON*** *Needed Number of* $NearbyNSs$

34 $Serve\ the\ IoTP\ Requests$

35 **end if**

36 **else**

37 ***Turn*** **ON** *Needed Number of* $NearbyNSs$

38 $Serve\ the\ IoTP\ Requests$

```
39          end if
40          Calculate Number of NSs & EnergySaving
41      end CASE
42
43      NRE with Non Energy Adaptive NSs + RDP CASE:
44      if IoTP Requests > 0 then
45
46          Find NearbyNSs
47          if Any NearbyNSs is ON & AvailableCapacity > 0
48              Calculate PDP
49              if RDP ≤ 0.05
50                  Serve the IoTP Requests
51              else
52                  Turn ON Needed Number of NearbyNSs
53                  Serve the IoTP Requests
54              end if
55          else
56              Turn ON Needed Number of NearbyNSs
57              Serve the IoTP Requests
58          end if
59      calculate EnergySaving & RDP
60      end CASE.3
61
```

Algorithm 10: IoT Information Piece Delivery with RDP

```
62      NRE with Energy Adaptive NSs + Max RDP CASE:
63      if IoTP Requests > 0 then
64
65          Find NearbyNSs
66          if Any NearbyNSs is ON & AvailableCapacity > 0
67              if IoTP Requests ≤ AvailableCapacity + Max PDP
68                  Serve the IoTP Requests
69              else
70                  Turn ON Needed Number of NearbyNSs
71                  Serve the IoTP Requests
72              end if
73          else
74              Turn ON Needed Number of NearbyNSs
75              Serve the IoTP Requests
76          end if
77      calculate Number of FSs & EnergySaving & PDP
78      end CASE
79      RE+NRE with Energy Adaptive NSs + OPT RDP CASE:
80      if IoTP Requests > 0 then
81          Find NearbyNSs
82          if Any NearbyNSs is ON
83              Calculate AvailableCapacity based on RE
```

Algorithm 10: IoT Information Piece Delivery with RDP

```
84          Calculate PDP
85          if RDP ≤ 0.05
86            Serve the IoTP Requests
87          else
88            Turn ON Needed Number of NearbyNSs
89            Serve the IoTP Requests
90          end if
91        else
92          Turn ON Needed Number of NearbyNSs
93          Serve the IoTP Requests
94        end if
95        calculate Number of NSs & EnergySaving & RDP
96      end CASE
97    end for
98    t1++
99  end for
```