Routing Strategies for Capacity Enhancement in Multi-hop Wireless Ad Hoc Networks

This thesis is submitted for the degree of Doctor of Philosophy (PhD)

Yu Ming Lu
Communication Research Group
Department of Electronics
University of York, UK

2007

Abstract

This thesis examines a Distributed Interference Impact Probing (DIIP) strategy for Wireless Ad hoc Networks (WANETs), using a novel cross-layer Minimum Impact Routing (MIR) protocol. Performance is judged in terms of interference reduction ratio, efficiency, and system and user capacity, which are calculated based on the measurement of Disturbed Nodes (DN). A large number of routing algorithms have been proposed with distinctive features aimed to overcome WANET's fundamental challenges, such as routing over a dynamic topology, scheduling broadcast signals using dynamic Media Access Control (MAC), and constraints on network scalability. However, the scalability problem of WANET cannot simply adapt the frequency reuse mechanism designed for traditional stationary cellular networks due to the relay burden, and there is no single comprehensive algorithm proposed for it.

DIIP enhances system and user capacity using a cross layer routing algorithm, MIR, using feedback from DIIP to balance transmit power in order to control hop length, which consequently changes the number of relays along the path. This maximizes the number of simultaneous transmitting nodes, and minimizes the interference impact, i.e. measured in terms of 'disturbed nodes'. The performance of MIR is examined compared with simple shortest-path routing. A WANET simulation model is configured to simulate both routing algorithms under multiple scenarios. The analysis has shown that once the transmitting range of a node changes, the total number of disturbed nodes along a path changes accordingly, hence the system and user capacity varies with interference impact variation. By carefully selecting a suitable link length, the neighbouring node density can be adjusted to reduce the total number of DN, and thereby allowing a higher spatial reuse ratio. In this case the system capacity can increase significantly as the number of nodes increases. In contrast, if the link length is chosen regardless of the negative impact of interference, capacity decreases. In addition, MIR diverts traffic from congested areas, such as the central part of a network or bottleneck points.

List of Contents

List of Tables

List of Figures

List of Publications

1. Yu Ming. Lu, "An Investigation of Routing Protocols for Broadband Wireless Communication Networks". Proceeding in IEEE PREP'02. Nottingham, UK. April 2002.
2. Yu Ming. Lu, "Resource Allocation Methods and Network Protocols for HeliNet" p.117~120 in EU Research Project-Framework 6-HeliNet-Task 4 internal Report HE-064. July 2002.
3. Yu Ming. Lu, "Routing Protocols for Next Generation Mobile Wireless Networks". Stage Report in Department of Electronics, University of York, UK. July 2003
4. Yu Ming. Lu, et al. "Performance Evaluation of Minimum Impact Routing for Multi-hop Wireless Ad Hoc Network". Proceeding in IEEE WPMC'04. Abano Terme, Padova, Italy. Sep. 2004.
5. David.Grace, Yu Ming. Lu, "NEWCOM: First Report on Common Models Matching Project A Needs - Cross-layer Routing Design: Coping with Link Asymmetry and Minimum Impact Routing" p. 37-40. in EU Research Network - NEWCOM. Oct. 2005.
6. Yu Ming. Lu, David. Grace, and Paul. D. Mitchell. "Capacity Evaluation of a Multi-hop Wireless Ad hoc Network Using Minimum Impact Routing". in IEEE WiCOM'06. Wu Han-China. Sep. 2006.

Acknowledgements

I would like to dedicate this thesis to my parents, as it is their continued support and encouragement that has enabled me to continually progress in my studies toward the completion of this thesis.

I would especially like to thank my two supervisors Dr David Grace and Dr Paul Mitchell for their continuous inspiration and guidance in my academic studies, and without their encouragement much of this work would not have been possible.

Special thanks go to Mr Tim Tozer, Prof Alister Burr, and Dr David Pearce, and Ms Dianne McLaren for the invaluable suggestions and discussions that helped and supported me through different stages of my study in York. I would like to thank everyone that I have known in the Communications Research Group through all these years, and although some of them have left before my submission of this thesis, still I thank them for the fun and joy they have given me.

Declaration

Some of the research in this thesis has resulted in publications in conference proceedings. All contributions presented in this thesis as original are as such to the best knowledge of the author. References and acknowledgements to other researchers in the field have been given as appropriate.

Chapter 1 Introduction

Contents

1.1 Background

This thesis investigates a Distributed Interference Impact Probing (DIIP) strategy and issues associated with routing protocol cross layer design for Wireless Ad hoc Networks (WANETs). A WANET is a self-configuring packet network of wireless routers and associated hosts (e.g. Personal Digital Assistants, Bluetooth Devices, Laptops etc.). Members of WANET can autonomously organize themselves into an arbitrary topology, and relay packets on behalf of other members, due to the assumption that not all members can directly communicate with each other[1-11].

WANET was initially developed for military applications. The earliest research by the US Department of Defence (DoD) can be traced back to the 1970s, under the project name of Packet Radio Network (PRN), which later evolved into the Survivable Adaptive Radio Networks (SURAN) programs in the 1980s [2]. The goal of the project PRN and SURAN was to provide packet switched communication networking to mobile elements in an infrastructureless or hostile environment such as soldiers, vehicles, ships, or aeroplanes [5, 6, 12, 13]. The unique and flexible characteristics of WANET, combined with the rapid evolution of electrical technology have given WANET great commercial potential. Following the success of cellular networks in the 1980s and IEEE 802.11 wireless LAN in 1990s, DoD continues the program under the name of Global Mobile Information Systems (GloMo), and Near-term Digital Radio (NTDR). The goal of project GloMo and NTDR was to provide office environments with multimedia connectivity anytime, anywhere. WANET provides low construction cost and a potentially unlimited wide range of applications, which have attracted an increasing number of researchers and developers to join in the force to accelerate the maturing process of pervasive deployment of WANET in modern society [2, 5, 6, 12, 13].

The peculiar characteristic of WANET is like a two-sided blade, one side shows great potential benefit for a wide range of applications, the other side produces tough challenges that prohibit the implementation of WANET.

The flexibility of WANET allows fast and easy deployment of such networks without the cost of base-stations and could serve users anywhere (i.e. in the air, sea, or on land) at anytime, in situations where the infrastructure is neither reliable nor available. This provides great potential for civilian (commercial or non-commercial) and military applications, such as the following examples [5, 6, 12, 13]:

- *Military autonomous networks* (e.g. personnel, vehicles, ships, and aircraft autonomous networks; sensors distributed in hazard areas etc.)
- *Environment monitoring systems* (e.g. sensors scattered in buildings located within earthquake or natural disaster high risk zones, pollution, wildlife, and environment change monitoring etc.)
- *Commercial WANETs* (e.g. ubiquitous computing for home and temporary offices; automobile networks; wireless gaming; extension of the internet etc.)
- *Public authority applications* (e.g. policing, traffic control, bailed criminal monitoring, disaster and emergency relief etc.);
- *Research and Scientific utilizations* (e.g. academic and research networks; undersea operations; space exploration etc.)

Before these beneficial applications turn into reality, many technical challenges lay ahead [5, 6, 12, 13]:

- The broadcast nature yields co-channel interference that limits both channel and system capacity.

- The lack of fixed infrastructure requires the network to deal with a dynamic topology, distributed control problems (e.g. Routing, MAC, Congestion Control, Administration etc.).
- The limited resources (e.g. spectrum, energy, channel capacity etc.) demand high consumption efficiency with fair distribution concerns (e.g. QoS).

WANET suffers a scalability problem due to its unique nature. WANET system performance is known to be limited, not only by the ***node level capacity***, e.g. raw *channel, node* or *link throughput* or capacity, but also by the ***network level capacity***, e.g. maximum *number of users* (or user population) supported by the system, or *network aggregated throughput* etc. Both means of capacity measurement are under the influence of factors of consequence, e.g. relaying burdens, interference (co-channel or adjacent-channel), energy constraint, node density, network size (or user population), traffic patterns, relay and topology variations etc [2, 6, 11, 13-44].

Scalability of such a network is associated with user population and the satisfiable channel capacity (i.e. how many users can a channel serve with an acceptable level of service). The more users that a system can support, the more scalable the network can be, e.g. in an omni-directional antenna environment where every node can reach the furthest node in one hop distance. The channel capacity of the throughput per node decreases, at a data rate of $\frac{1}{\sqrt{N}}$, where N is the number of nodes (or user population). If a network has 100 nodes, each node can only get approximately one tenth (1/10) of the theoretical maximum data rate [8]. This is due to multiple impact from relaying burdens, interference etc. We refer to the effect where a transmitting node interferes with surrounding co-channel nodes while communicating, and the resulting effect on system capacity, as the "interference impact"[17, 33, 36, 37].

Previous research has attempted to relate and resolve this problem via many means, including routing protocol design. The *proactive* (i.e. constantly maintained by a global topology in each routing table) type of routing protocol cannot cope with the dynamic network topology, so that the routing control overheads will overflow the whole network. The *reactive* (i.e. discover and maintain a path in an on-demand manner) type of protocol will allow the user population (i.e. number of nodes) to be increased at the expense of proportionally increased route acquisition latency. *Hierarchical* routing (i.e. operating routing and other network functions on several hierarchical levels) can relieve the scalability problems to a certain degree via clustering, i.e. proactively routing within a cluster, and reactively routing outside the cluster. It is one of the few methods that can cut down the proportion of overheads, and shorten the routing acquisition latency.

However the growing demand on channel capacity and the user population capacity cannot be satisfied by any single technique, but instead, a combination of effects of multiple techniques across the functional layers of a system, as listed below:

- For the *physical layer*, dynamic channel characteristic changes channel and physical devices adaptations.
- For the *Multiple Access Control* (MAC) layer, distributed scheduling takes on the task of minimizing collisions for fair access, avoiding hidden-terminal transmissions.
- For the *network* layer, dynamic routing distributes information to discover and maintain connectivity of paths between nodes, whilst interconnecting with conventional systems.
- For the *transport* layer, distributed traffic congestion, packet loss, delay, and retransmission control manages packet or stream transmissions.
- For the *application* layer, distributed disconnection and reconnection management with peer-to-peer applications.

Nevertheless the complexity of the wireless environment is in conflict with the growing demand for capacity and new applications from the ambitioned visionaries. The question of whether WANET can provide an acceptable level of channel capacity, even in the presence of a large number of nodes in the network, and furthermore, how large can WANET grow has become an ever more challenging task.

1.2 Scope of this research

This thesis examines research carried out on a Distributed Interference Impact Probing (DIIP) strategy, based on a cross layer routing protocol that is designed and developed with an aggregation of techniques, for partially overcoming the scalability problem in multi-hop WANET.

The first objective of this research is to develop a deep and thorough understanding of the technologies and issues associated with the interference, relaying, overhead impact, protocol design (i.e. routing, MAC, physical layer sensing etc.) and system scalability, which has resulted in development of a unique Distributed Interference Impact Probing (DIIP) architecture through a comprehensive literature review.

The follow on work is focused on designing, developing and improving the Minimum Impact Routing (MIR) protocol, which uses an original approach of DIIP that is dedicated to minimizing the co-channel interference impact, and utilises the outcome of DIIP as a routing criteria. MIR tackles the scalability problem from a network layer perspective, optimizing the spatial reuse, and therefore enhancing he WANET system capacity and user population in terms of maximizing the total number of simultaneous transmitting nodes in the network. There is a more detailed discussion and proof of this in later chapters.

MIR is improved with a range of dynamic techniques to enhance its adaptability in WANET. The following improvements are made throughout each stage of research:

- Stage one is to modify the primitive reactive MIR routing to a hierarchical type of routing operation. The major changes include: adaptation of *Next Forwarding Nodes* (NFN) for a *Controlled Flooding*; *Local Communication Group* (LCG) clustering for *Dynamic Topology Control* (DTC). This resulted in the original adaptive MIR.
- Stage two is the integration of an adaptive MAC protocol, Busy Tone Multiple Access, which is particularly beneficial in providing a reduction in control traffic (in other words with no Request and Acknowledgement packets), and solving hidden terminal problems during the transmission.
- Stage three is to integrate the Variable Transmit Power in connection with the DIIP measure, which is a unique and original contribution, and allows wireless nodes to dynamically adjust their transmit power based on the surrounding node

density while forming the LCG. This resulted in a MIR-VTP protocol published at an international conference, WPMC, in sep 2004 [17, 33, 36, 37].

- Stage four is the further upgrading of the MIR-VTP (i.e. MIR with Variable-Transmit Power) to operate in a unidirectional environment where asymmetrical links are studied and simulated. The result of this study was the novel MIR-VA (i.e. MIR with Variable-transmit-power plus Asymmetrical-routing) protocol, which uses a unique asymmetrical transmitting of the routing control packets to achieve the asymmetrical routing. This resulted in a contribution to the routing model in a European research project called NEWCOM, and another international publication in WiCOM Oct 2006, which describes the spatial reuse theory based on a concept called Time Sequenced Interference Region (TSIR).

The final analysis of the complex simulation results, collected from those simulation models composed in early stages, had an inspiring consequence in deriving the findings conclude later. It triggered the derivation and development of an original WANET system capacity model, and a new series of innovative WANET system capacity enhancement strategies.

1.3 Structure of the Thesis

This thesis explores different aspects associated with research conducted on reducing the scalability problem in WANET, using a DIIP mechanism. This research results in a new family of MIR metrics, which incorporate DIIP, LCG, variable transmit power, and asymmetrical routing. The MIR routing algorithm is a cross layer aggregation developed in order to achieve high system user population, and is extended from previous mathematical analysis [3, 16, 18, 20, 21, 45].

Chapter 2 reviews the challenging issues and technologies that are essentially associated with effective routing strategies which are aimed at enhancing the system capacity and reducing or eliminating the scalability problem in WANET. Some influential factors such as WANET features and challenges; transmit power control; multiple access techniques; routing challenges; and system level considerations are included.

Chapter 3 provides an overview of issues associated with WANET routing. A comprehensive literature review of related areas is presented, in terms of routing environment and routing algorithms, in order to have a deep and thorough understanding of technologies and issues associated with interference environment and impact, protocol design (i.e. routing, MAC, physical layer sensing etc.), system capacity, and related techniques, used in the development of a unique Distributed Interference Impact Probing (DIIP) architecture, in which these factors are incorporated by a novel routing algorithm, Minimum Impact Routing (MIR), aimed at enhancing system scalability in WANET.

Chapter 4 introduces the modelling methodology, statistical result collection and measurement, and the validation and evaluation methodology. Among these are a brief review of the simulation tool (i.e. OPNET), the different means of design, simulation, result collecting, and result evaluation.

Chapter 5 is dedicated to the analysis of system performance, and identifies the interconnections of multiple influencing factors in system operation. This analysis is lead by the three important trade-off relationships. Some analysis is based on the abstractive assumptions to simplify the situation, and this is supported with a series MATLAB simulation results.

Chapter 6 presents a description of a novel MIR protocol, as well as some related aspects. The interference impact is firstly quantified using a Disturbed Node (DN) concept, then a DIIP mechanism is developed to measure such impact using DNs, which reflect the local interference impact, and define a Local Communication Group (LCG) with a constrained interference impact region. The routing decision is made using a measure of accumulated DNs along a path, as the criteria to calculate a shortest path between two nodes. In order to study the performance of MIR, a 32-node simulation model is developed, with three basic topology scenario. The simulation results are analysed from a network layer perspective, and verified. This stage of development resulted in an international publication.

Chapter 7 introduces the later improvement of MIR, the more mature MIR-VTP protocol, with details of essential theories and techniques such as: variable transmit power, spatial reuse theory, asymmetrical routing strategy, path capacity derivation based on the time sequenced interference impact region concept, and consequently the network user capacity derivation. This section also provides a description of a series of simulations conducted using an improved model, under different network topologies and scenario configurations, using different routing algorithms. The resulting analysis reviews the adaptive routing algorithm such as MIR-VA, which incorporates the interference impact evaluation and balancing into the routing operation, resulting in increased system scalability, under optimal local interference impact threshold that is measured in disturbed nodes. In which case, this extended the capacity boundary defined by Gupta and Kumar in [20], and the mathematical analysis of capacity region by Toumpis and Goldsmith, uses variable transmit power and interference cancellation. The theory of MIR-VA is then verified using simulation result analysis and evaluation. This stage of research resulted in another international publication.

Further more, a 60-node WANET model is developed using OPNET in order to simulate and evaluate the centralized or distributed type of network control scenario, with the variation of single-hop or multihop transmissions, with or without spatial reuse, based on a random topology.

Chapter 8 presents the possible future extension of the current research. This includes the improvement of the current unified channel capacity assumption; diversity of interference impact measuring; improvement of distributed MAC schemes; extended study of the complex network scenarios, such as multiple random network coexistence; and the extended capacity analysis for Frequency, Time, and Code Division Multiple Access (F/T/CDMA) systems.

This is followed by the overall summary and conclusion in chapter 9.

Chapter 2 Literature Review

Contents

2.1 Introduction

This chapter presents a literature review of the key issues that influence the system capacity and behaviour of a Wireless Ad Hoc Network (WANET). The distinctive features of WANET provoke the technological challenges in physical characteristics management, power control, contention control, routing, spatial reuse, system level, and other higher layer function design.

Early research on WANET can be traced back to the 1970s, when it focused on providing low data rate packet switched wireless communication networking for military applications. It was only after the cellular systems successfully delivered *high* *system scalability*, in terms of user populations of a system, with a frequency reuse technique implemented in cellular structures in the 1980s; and the Wireless LAN (WLAN) demonstrated the feasibility of Ethernet type *high channel capacity* (i.e. high data rate assigned to each user) for mobile computing in the 1990s, that global interest in WANET re-emerged. The technological development of cellular networks, WLAN, and Bluetooth, enables much higher data rate communication from small wireless terminals, such as mobile phones, laptops and Bluetooth devices, to support a large amount of users with centralized control mechanisms.

The question that remains unsolved is: Can we develop a high channel and system capacity for WANET by adapting and improving technologies utilized in conventional wireless communication networks, such as cellular, WLAN, Bluetooth etc, to support infrastructureless high data rate mobile communications with high user population? In terms of channel capacity, WLAN can provide higher data rates in contrast to cellular networks, due to the use of high transmission frequency, and relying on the centralized control at a stationary access point that is connected to the fixed infrastructure.

However WLAN cannot support the same user population as a cellular network, due to the uncoordinated frequency reuse planning between WLAN access points. On the contrary, cellular networks can support high system user populations with the frequency reuse mechanism carefully organized between stationary base stations. This chapter discusses the interrelationship of key factors and technologies for conventional wireless systems, later discussed in detail in chapter 5 in order to understand their impact on solving the *scalability problem* in WANET, and proposing an innovative routing mechanism for WANET.

2.2 Wireless Ad Hoc Networking

The trend of future communication networks is to combine conventional cable or wireless systems with Internet or future generation networks (e.g. 4G or Universal Telecommunication Systems UMTS). The major difference between WANET and other wireless systems is the capability of operating independently without the support of a fixed infrastructure. Accelerating WANET research in recent years has generated numerous potential application proposals, and consequently more challenges are realized. This section outlines the advantages and potential applications, together with the disadvantages and challenges of WANET.

According to the coverage areas, modern wireless communication systems can be broadly divided into four categories[12]:

- Wireless Body Area Networks (WBANs), i.e. wearable devices or components distributed on a body, interconnected using Bluetooth or infrared technology.
- Wireless Personal Area Networks (WPANs), i.e. wireless devices carried by a person interconnected with other mobile or stationary devices in the environment around a person.
- Wireless Local Area Networks (WLANs), i.e. wireless mobile or stationary devices interconnected in home or office with a communication range of, in a single or a cluster of buildings, up to 500 meters.
- Wireless Wide Area Networks (WWANs), i.e. widely distributed wireless devices across an area in the order of kilometres, such as urban areas, or villages.

Figure 1 shows the coverage region, interconnection and comparison between cable-based modern wireless and future integrated systems. In terms of existing or under development standards utilizing these networks, WBAN and WPAN are small area networks for wearable computers or Bluetooth devices, using communication standards such as Bluetooth (IEEE 802.15); WLAN is the most popular for mobile computing using IEEE 802.11 (also know as WiFi) standards. WWAN covers larger areas and is aimed at using standards under development such as IEEE 802.15.5 (Mesh network), IEEE 802.16 (also known as WiMAX) for urban areas, the countryside, and even between countries.

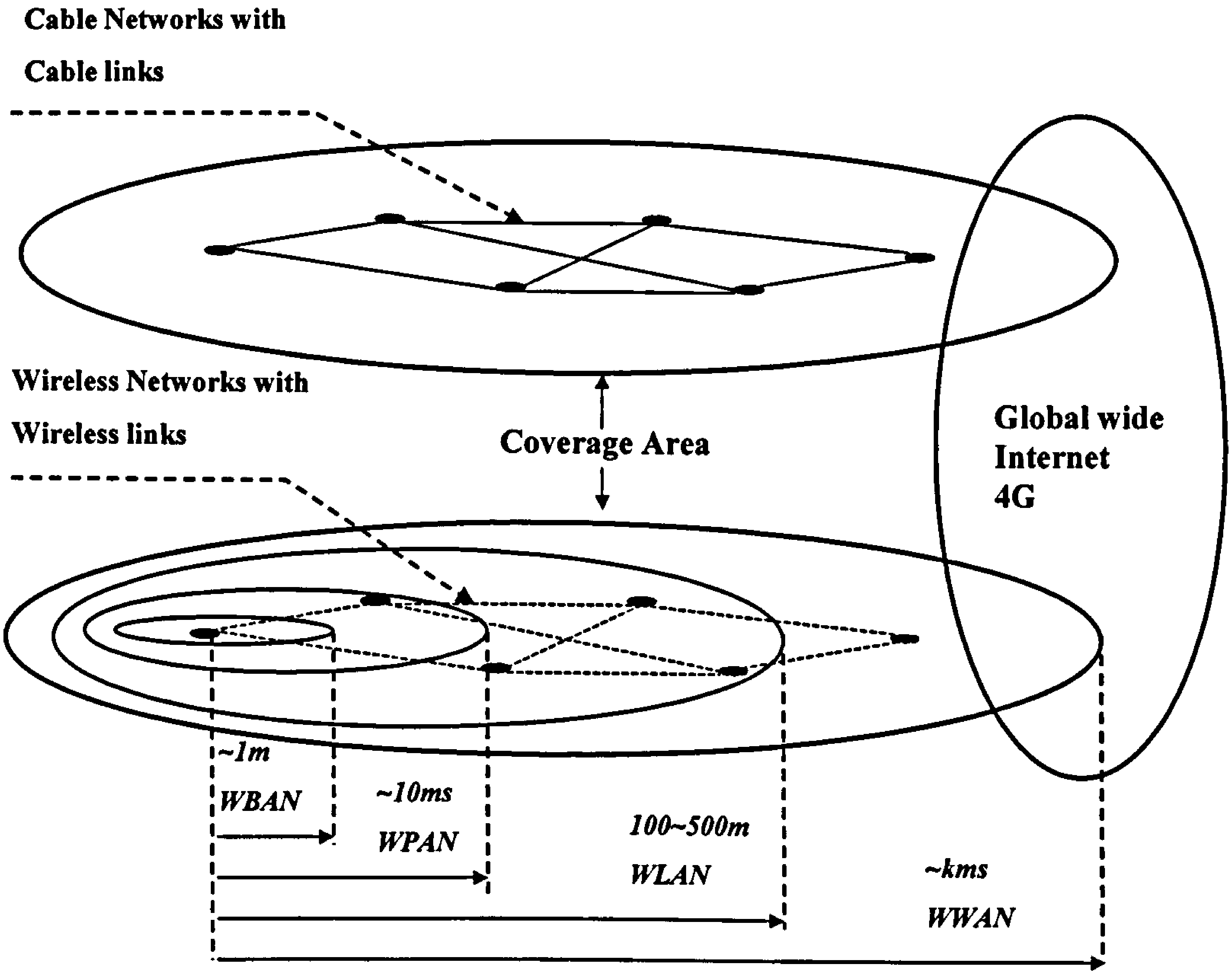


Figure 1 Type of Network and Coverage [12]

2.2.1 Centralized vs. Distributed Control

The distributed control is one of the fundamental differences between WANET and other wireless communication systems that use centralized controls. As a conceptual network, WANET provide means of use that can be deployed in any type of conventional networks (i.e. WBAN, WPAN, WLAN, and WWAN). However the technologies used in WANET are *distributed type* controls, whereas existing system standards for WBAN, WPAN, WLAN, and WWAN are more or less based on *centralized controls.*

Centralized control operations are conducted and coordinated by a central location, e.g. base stations of a cellular or early radio network; an access point of Wireless LAN, or master devices of Bluetooth *piconets*.

Distributed control operations on the contrary, are used by the user terminals, i.e. mobile or stationary wireless terminals or devices. This gives WANET great advantages, such as highly flexible structure and configuration, fast and easy deployment, and low cost implementation etc., for potential applications. However these existing application opportunities also pose significant technical challenges, which involve cross layer design and finite-capacity resource (e.g. spectrum, spatial, and energy etc) management.

2.2.2 Advantages and Applications of WANET

The visionary pioneers proposed applications of WANET covering a wide range of aspects in our society associated with daily life. Without the burden of fixed infrastructure, WANET could overcome many limitations that are inhibited in conventional communication networks. *Military* applications initiated the WANET research in order to design and implement autonomous networks that interconnect personnel, battle vehicles, ships, and aircraft distributed in the battlefield or operating areas; sensors deployed in hazardous or monitoring regions for information gathering. *Civilian* applications of WANETs cover a wide range of daily utilizations, for example:

- *Environment monitoring systems*, i.e. sensors scattered in buildings located within earthquake or regions at high risk of hazard and natural disaster; nuclear, industrial or chemically polluted areas monitoring; wildlife research tracking and monitoring; constantly monitoring climate, temperature, and other environmental changes etc.
- *Commercial WANETs*, i.e. ubiquitous computing for the home and temporary offices; automobile or traffic networks along the motorways; indoor or outdoor wireless gaming; wireless extension of the internet; industrial property tracking;

farming; storage; retail item tagging; intelligent buildings; medical networks etc.

- *Public authority applications*, i.e. police patrol; home and building security monitoring; traffic control; bailed criminal monitoring; crime watch; facility usage statistics; disaster and emergency relief etc.
- *Research and scientific utilizations*, i.e. educational networks for academic study, medical research, remote teaching, and outdoor scientific research; undersea operations and investigations; space exploration etc.

WANET provides anytime and anywhere applications under circumstances such as: where a cable network is either impractical or not reliable; locations that are beyond reach; environments where it is dangerous for the presence of humans; highly random network structures etc. The numerous potential applications have put WANET in a vital position in future integrated networks.

2.2.3 Disadvantages and Challenges of WANET

Researchers must overcome many challenges before the wide deployment of WANET is a reality, due to the three inherent disadvantageous features of WANET:

- the broadcast nature of communication in complex wireless environment;
- the lack of support of the fixed infrastructure;
- the limited resource;

Table 1 shows the broadly summarised interconnection between *features of WANET*, the challenges that these features pose, and the related issues in relation to the different functional layers.

WANET Features	Posed Challenges and Trade-Offs				Routing related Tasks	Routing related Methods
Broadcast Nature	Co-Channel & adjacent channel interference (connectivity & Spatial Reuse); Multiple Access contention and collisions	Relaying Delays, Controls Overhead (Single Hop versus Multi-hop Routing)	Interference-Based Controls and Multiple Access Control (Centralized versus Distributed)	Connectivity Scalability and Spatial Reuse (Connectivity versus Scalability)	(PHY): Achieve adequate SNR	(PHY): Distributed Power Control (DPC); Distributed Link Quality Control (QCL)
					(MAC): Contention and Collision control; Multiple Access control	(MAC): Contention-Based (e.g. BTMA; CSMA); Contention-Free (e.g. Polling)
Lack of Infrastructure **(Flexibility in Link Connection)**	Distributed Networking and Control Functions				(PHY): Distributed Interference Control	(PHY): Distributed Interference Control use Variable Transmit Power (e.g. DPC)
	Dynamic Topology				(MAC): Distributed Multiple Access Control	(MAC): Distributed Carrier Sensing Multiple Access Control (D-CSMA)
	Asymmetrical Connections				(NET): Distributed Topology Control: Distributed Routing control over Asymmetrical links;	(NET): Distributed Topology Control (e.g. Periodical update packets; Clustering etc); Distributed Routing (e.g. Proactive, Reactive, Hierarchical, and Asymmetrical routing etc)
Restrained Resource	Spatial Efficiency (Scalability)				(Cross-Layer): Spatial Reuse	(PHY +MAC+ NET): DPC + D-MAC + D-Routing
	Energy Efficiency (Power Conservation)				(Cross-Layer): Power Conservation (not discussed in detail in this thesis)	(PHY): Distributed Power Control (DPC): Distributed Link Quality Control (QCL)
	Spectrum Efficiency (Scalability)				(Cross Layer): Spectrum Management (not discussed in detail in this thesis)	(Future work): Spectrum Sensing

Table 1 Overview of the Cross Layer Design Challenges, Tasks and Suggested Method in relation to functional layers (more detail in chapter 5)

2.2.3.1 Broadcast Nature of WANET

WANET inherited the broadcast nature of conventional wireless networks. The highly complex wireless environment creates adjacent and co-channel ***interference*** demanding efficient *multiple access and contention control*, and the restricted spectrum and spatial resources together concerns WANET development deeply. In such an environment, both the *adjacent channel* (i.e. neighbouring frequency channels used by different users) and the *co-channel* (i.e. the same frequency channel used by different users) interference will result in negative effects (e.g. noise and collisions) to the communication. Consequently this results in a reduction of the channel capacity for each user, and reduces system user population for the whole network. Interference from a co-channel node in the same region may cause transmission collisions that eventually make wireless communication impractical [1, 12, 13, 42].

In terms of ***spatial reuse***, the interfering nature of broadcasting wireless systems needs careful spacing of interference sources to make sure the two co-channel transmitting nodes are spatially separated, so that they do not become a source of interference for each other. In cellular networks spatial reuse is often implemented by using a frequency reuse mechanism, which carefully spaces the co-channel cells in a cellular structure[27, 42]. In WLAN a master node can use a frequency hopping spread spectrum (FHSS) technique to poll all wireless users (slaves), which employs contention-free services inside each single cell, in this case spatially separating the co-channel interference[1, 11-14]. WLAN also supports contention-based communications using Direct Sequence Spread Spectrum (DSSS, which is also known as CDMA), however this type of transmission poses side effects problems such as the *Near-far* problem also known as the *capture effect* [13, 46-49]. This requires a distributed power control scheme to prevent the stronger transmitter from capturing a receiver whilst other weaker transmitters are also trying to communicate with the same receiver. This problem will be discussed in further detail in the next section.

In a contention-based wireless network, without an efficient ***contention control*** for *collision detection* and a *spatial reuse* mechanism, the co-channel interference can seriously damage WANET system and user capacity. In conventional wireless systems, such as cellular networks and WLAN, contention control was implemented using Multiple Access Control (MAC) protocols with a collision detection function called

Carrier Sense Multiple Access with a Collision Detection / Avoidance (CSMA-CD/CA) [42, 43].

2.2.3.2 Lack of Infrastructure

The lack of support of a fixed infrastructure in WANET gives it flexibility for deployment and mobility, but it also poses challenges such as ***distributed controls*** and ***dynamic topology***. There is no centralized control from a base station or similar central locations, which means the burden of conventional centralized control (e.g. routing, flow, congestion, contention, channel characteristics, transmit power control etc.) of a base-station, router or bridge, now has to be shared by end-users in a distributed manner. In terms of *routing*, the fast changing network configuration due to the node movement, and nodes joining and leaving the network, yields changes in the network topology. These topology changes trigger frequent route or topology updates that generate massive route control overheads, which have a negative effect on network traffic. Also because of the lack of central control, members of WANET have to cooperate with each other to coordinate and relay traffic, and allocate valuable resources etc.

2.2.3.3 Restricted Resources

WANET has limited resources, such as bandwidth and energy. Members of WANET are most likely to be battery powered wireless terminals, which means they have a limited power source to operate in a limited lifetime. ***Energy and spectrum efficiency*** are two essential aspects that need to be considered in a WANET system and protocol design, so that the system can carry out energy and spectrum efficient operations. A variety of methods have been proposed for energy conservation and spectrum efficiency. In system design, the total energy consumption has been concluded as the sum of computation and communication energy consumption [10].

Numerous power control, modulation and coding, and quality of service mechanisms have been proposed for high efficiency system design[1, 2, 4, 10, 25, 26, 38, 40, 41, 50-59]. In protocol design, power aware and spectrum efficient routing and MAC protocols are proposed with the consideration of energy conservation and spectrum utilization efficiency [2, 4, 8, 9, 52-54, 60]. The WANET has an unique characteristics of requires nodes to cooperate with each other, and relay traffic on behalf of others.

This unavoidable duty of relaying consumes both energy and spectrum in the network, therefore the nature of relaying needs to be carefully studied and organized to ensure fair share of relay burden, and diverting traffic to prevent congestion is essential.

In order to analyse WANET performance, one needs to abstract out the essential aspects that governs the performance of the system. This is a complex and challenging task since such analysis must take into account the interactions between challenging problems that are associated with the three essential system functional layers, which are broadly classified as *the Physical layer, the Media Access Control (MAC) layer,* and *the Network layer*. The *Physical layer* deals with channel characteristics such as the use of *variable transmit power* to adjust signal strength to ensure appropriate *Signal to Interference and Noise Ratio (SINR).* The MAC layer takes care of *contention controls* during the communication. The network layer handles *addressing* and *routing*, which ensures the network *connectivity* and *topology control.*

2.3 Physical Layer

In telecommunication, the FM ***capture effect*** is a phenomenon associated with FM reception in which only the stronger of two signals at, or near, the same frequency will be demodulated, when the weaker signal at the receiver is not amplified, but attenuated and suppressed. For a nearly equal in strength, and independently fading signal, the FM transmitter will cut in and out as it nears the capture threshold of the receiver, this is so called *picket fencing*. Therefore some applications chose to use AM radio instead, since in digital modulation schemes, it has been shown that OOK/ASK systems are more capable in co-channel rejection than FSK systems.[42, 43]

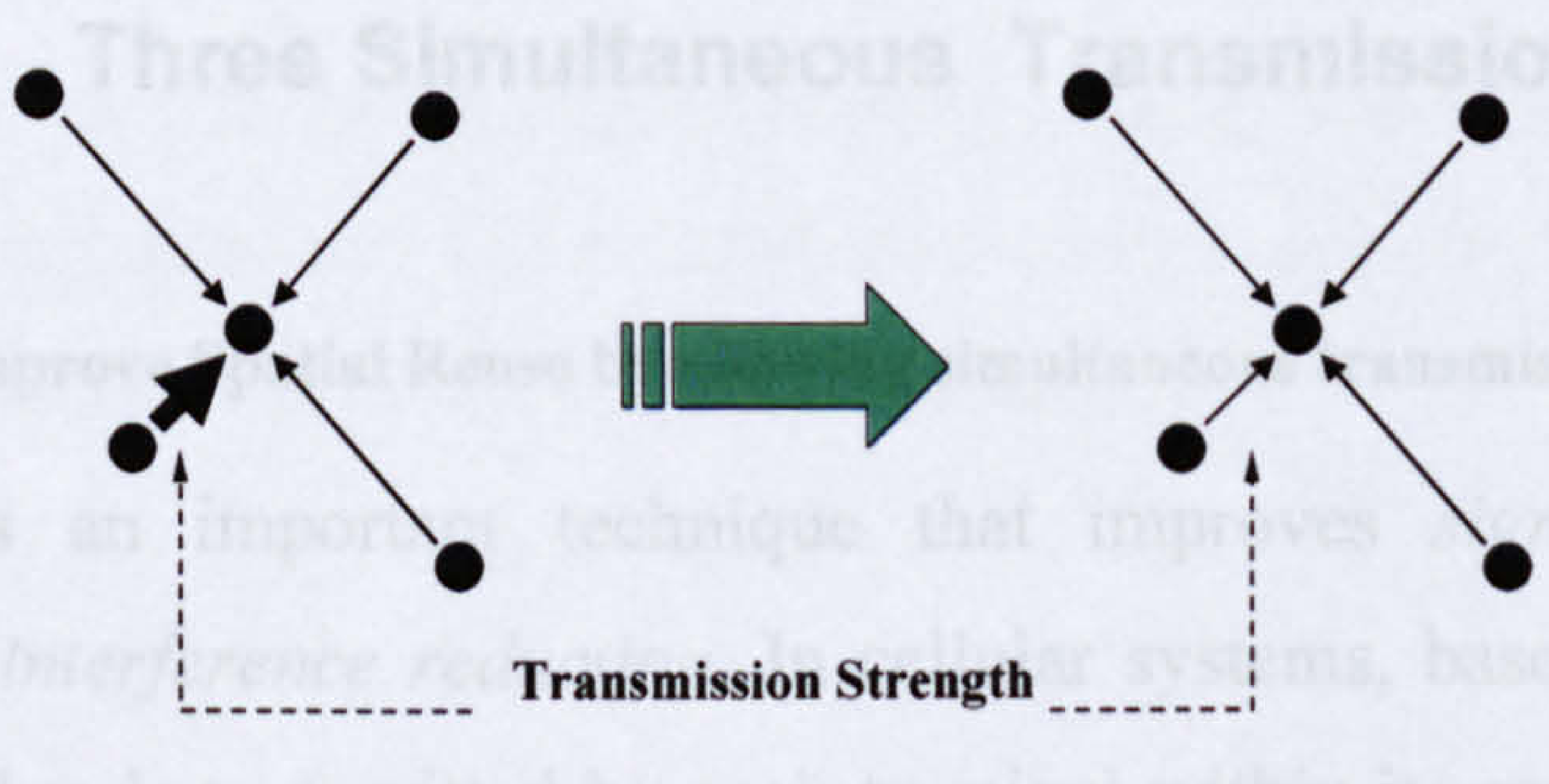
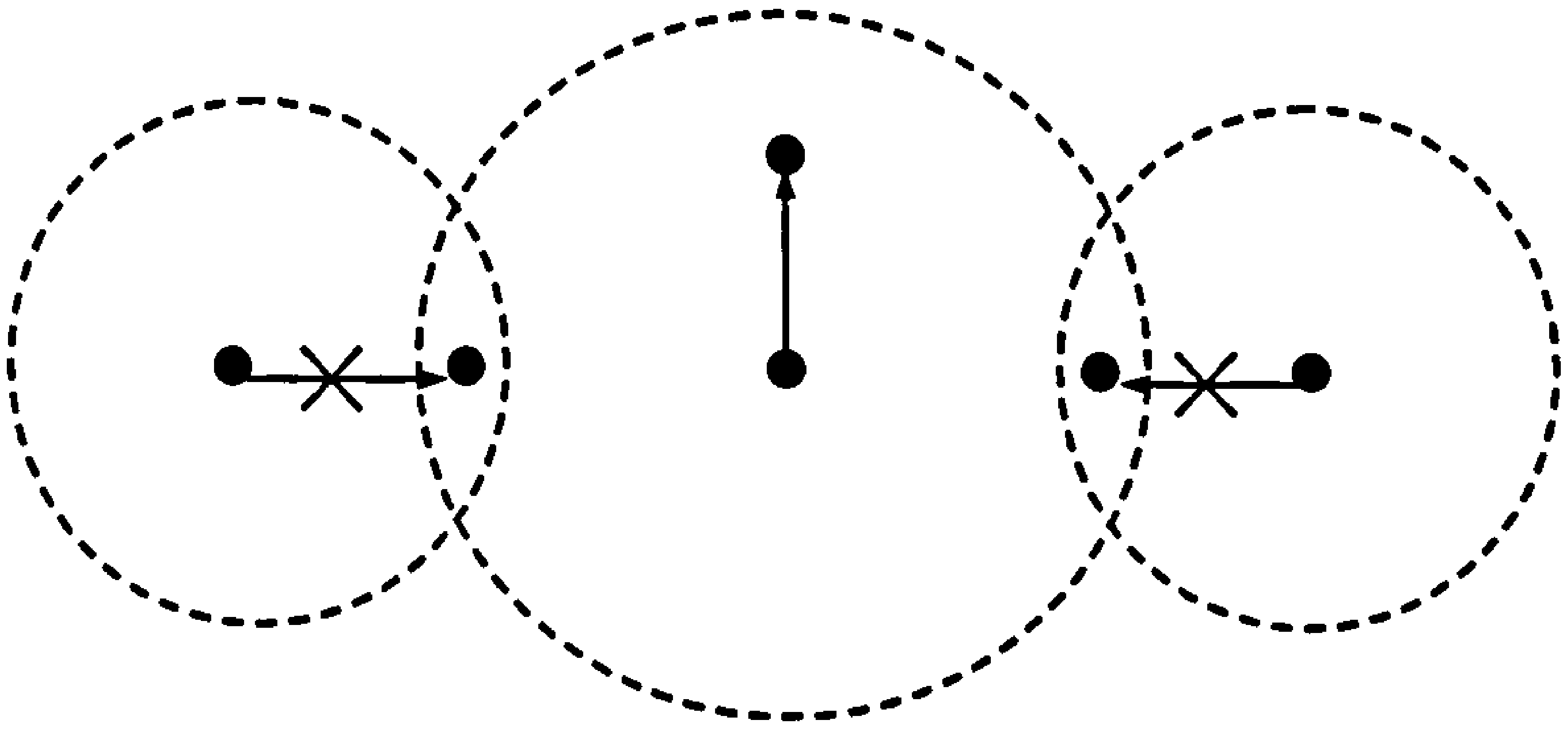
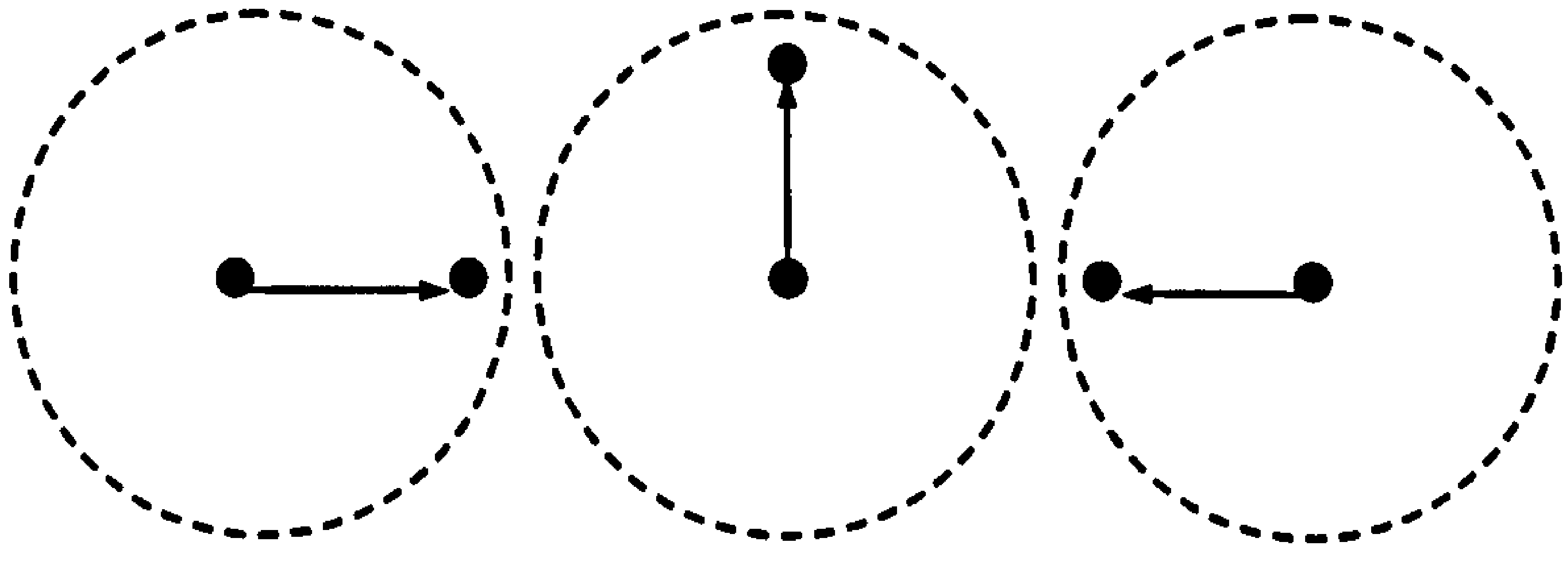


Figure 2 Near- Far problem [13, 42-44]

The other devastating effect in wireless communication systems is the ***Near Far***, or ***Hearability*** problem. In CDMA systems, if two simultaneous transmitting nodes (i.e. one nearer, one farther) use equal transmit power that shares a common receiver, then the receiver will receive, due to inverse square law, higher power from the nearer transmitter than the farther one. The Signal-to-Noise-Ratio (SNR) for the farther transmitter is much lower so its signal may not be detectable by the receiver, hence it may as well not to transmit. This has effectively closed the communication channel. This problem is commonly solved by exhibiting physical layer *Power Control* i.e. dynamically adjusting the transmitter's output power, so that the closer transmitter uses less power the SNR from both transmitters is roughly the same at the receiver. Sometimes this has a significant impact on prolonging battery life. Figure 2 illustrates the principle of this phenomenon.



One Concurrent Transmission



Three Simultaneous Transmissions

Figure 3 Improve Spatial Reuse by allowing simultaneous transmission [42, 43]

Power Control is an important technique that improves *signal quality*, *energy conservation*, and *interference reduction*. In cellular systems, base-stations constantly control the power levels transmitted by each terminal within its service coverage area, so that each terminal transmits at the smallest power level necessary to maintain usable SNR or signal quality, prolonging their battery life, and at the same time reducing the interference between transmitting nodes to an acceptable level. Base stations rapidly sample the radio signal strength level from each user terminal. If a user's SNR level is above a threshold, i.e. signal strength too high, it will reduce its transmitting power level.

In distributed systems, power control also provides high spatial reuse efficiency by allowing more simultaneous transmissions. Figure 3 demonstrates a distributed system power control, in which the reduction of transmit power level enables multiple simultaneous transmissions in the network.

The **drawback** of power control is the variation in wireless links may result in *weakened connectivity*, due to reduced signal strength, and the *Power Control Run Away* situation. The Power Control Run Away is a process which occurs when the nearer transmitter raises output power to improve its SNR in a high-noise situation, in which case it forces the farther transmitter to raise output to maintain good SNR. Other neighbours react to this raising noise floor by increasing their transmit power accordingly, since the signal of one transmitter is noise to the others. Eventually the farther transmitter unable to match the increasing noise floor and maintain a usable SNR, drops out from the network. This principle explains why the service quality of a system could degrades significantly when the traffic load increase.

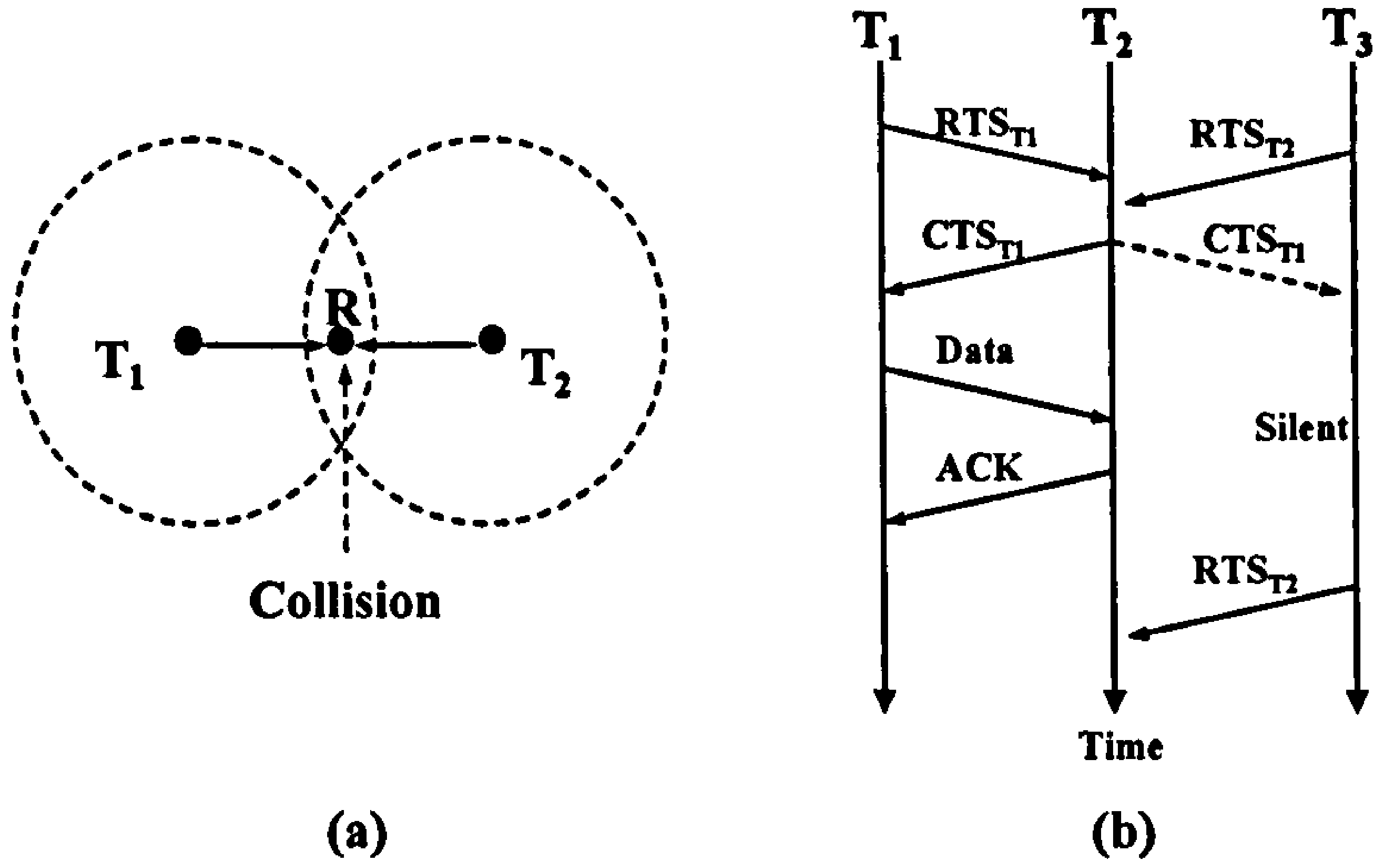
The power control, link quality variation, dynamic network topology, control overheads, and traffic intensity are important factors for the performance of a wireless system. In some cases it is arguable that using a longer or shorter link length will be more beneficial to connectivity, transmission success rate, less relays, energy conservation, and overheads reduction [10]. However short link length may improve spatial reuse by allowing more simultaneous transmission, which is a desirable feature in terms of supporting more users in a distributed wireless system like WANET. These issues are further discussed in chapter 5 in more detail.

2.4 MAC Layer

Contention control is the main function of Media Access Control (MAC) layer in the traditional Open System Interconnection (OSI) reference model. In a contention-based communication network, transmitters of a user terminal compete with other co-channel nodes. In this process, without any contention control, simultaneous transmissions competing for the same receiver will most likely result in collisions at the receiving end. This is also known as the *hidden terminal* problem, which in conventional centralised wireless systems, is dealt with by the MAC scheme, such as Carrier Sense Multiple Access with a Collision Detection / Collision Avoidance (CSMA-CD/CA).

Figure 4 (a) demonstrates the scenario where both transmitters, T_1 and T_3, are trying to communicate with the same receiver T_2, but remain hidden from each other. The simultaneous transmission for these two transmitters will collide at R, as the result of lack of contention control.

Figure 4 (b) shows the CSMA uses Request To Send (RTS) and Clear To Send (CTS) dialogue to organize the transmission sequence of the pair of competing transmitters. When T_3 overhears the CTS message for T_1, with the time required for the communication, T_3 will remain silent for the period of time that T_1 and T_2 are communicating, and retry its RTS message after the communication[42, 43].



(a)

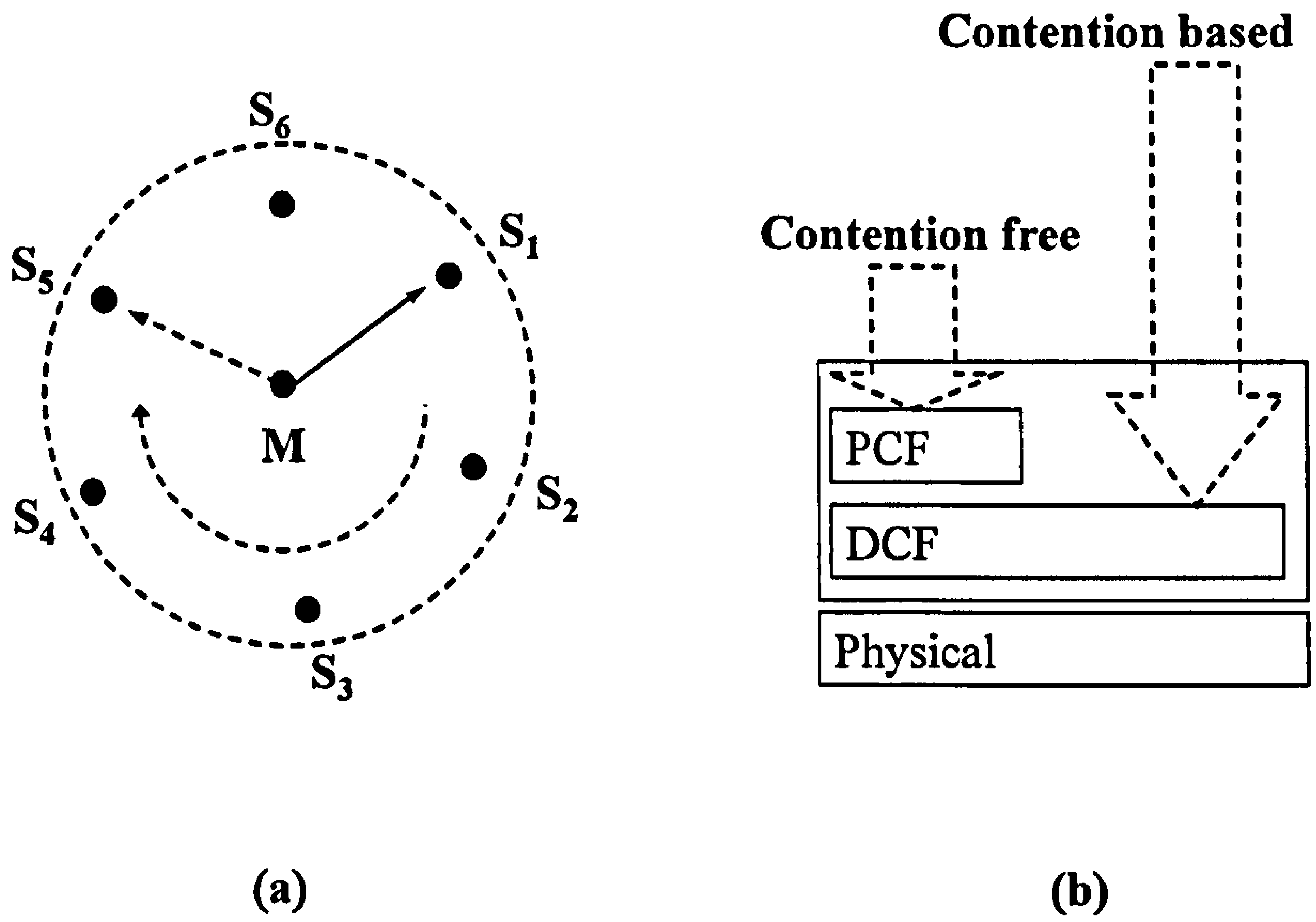
(b)

Figure 4 Hidden Terminal Problem

(a) Hidden Terminals T1 and T3 Collide at T2 (b) RTS-CTS dialogue between T1 and T2

Another way of solving the collision problem is to use a contention-free multiple access scheme, called polling[43, 44, 46, 51, 61-65]. In this case, wireless terminals get regular questioning from the central controller, asking if they have any data to send. The central controller then polls in sequence all wireless terminals within its coverage area. The IEEE 802.15 Bluetooth standard uses this mechanism as its contention control.

A master node polls all slave nodes within the same *Piconet* cell, and allows multiple access to take place in turns. IEEE 802.11 WLAN uses a Point Coordination Function (PCF), which is similar to a polling system, for the contention-free traffic, and a Distributed Coordination Function (DCF), which is a CSMA-CA MAC protocol for contention-based type of traffic coming through each access point [12, 13, 42].



(a) (b)

Figure 5 Multiple Access Control Schemes

(a) Polling in Bluetooth (b) IEEE 802.11 WLAN Point Coordination Function -PCF and Distributed Coordination Function-DCF

Figure 5 (a) shows a simple polling mechanism, where (b) shows the MAC layer mechanism of IEEE 802.11 WLAN. Both of above described methods can effectively solve the collision problem in a contention-based wireless, however WANET cannot directly adapt these conventional contention control schemes due to their lack of centralised control nature. Therefore further study on the conversion of contention control schemes is an important issue for WANET implementation.

2.5 Network Layer

Routing is the premier task of the network layer for determining the best route between two points in a network, and is the most fundamental research issue in WANET. The main function of routing is to learn existing destinations, and establish connections between source and destination for later relaying or initiating data traffic. The distinctive feature of WANET has redefined the characteristics for routing algorithm and protocol design.

The fundamental difference between conventional routing and routing in WANET is that conventional system uses *centralised routing* operating in a stationary central location, such as routers or base-stations, whereas WANET routing operates in a *distributed* manner carried out by each user terminal, which is possibly mobile [1, 2, 12, 13, 42]. Stale or duplicated packets in the network could cause the network traffic overflow. A sequence number in conjunction with a packet lifetime could be used to prevent the stale packet travelling in the network being duplicated endlessly [66]. Table 2 shows the broadly classified types of routing that appear in sequence.

Conventional **centralized routing** collects and maintains routing information (e.g. topology, route cost, updates etc) at a centralized location (e.g. router, hub, or a base station), and interconnected these stations to form a larger network. The routing information updates are broadcast to all these stations, so that each station will determine one (or multiple) best path to reach other remote stations, and record this path in the entries of its own routing table. The routing table is usually constructed using a routing matrix, which consists of a row index and a column index that represent the known source and destination nodes accordingly. Each source node is indexed by a row of destination nodes, each has an entry, which records a next/first relay node towards that destination.

Routing Types	Advantages	Disadvantages
Centralised Routing	***Static routing***: Dijkstra's /Shortest path: simple method use unchanged / estimated routing information. Hence low latency. Simple static routing tables with low latency.	Insensitive to the network condition changes. Failure on central stations or any links result in severe network breakdowns.
	Adaptive routing: Bellman Ford/Distance Vector: Routing decision made reflects network changes dynamically.	High latency, high overhead.
Distributed Routing	***Proactive***: lower latency since routing information are proactively updated	high routing overhead for maintain routing information update.
	Reactive: low routing overhead due to the less frequent routing acquisition from the source, which also means higher spectrum efficiency.	High latency caused by node-by-node update propagations. Poor scalability since overhead mount up as the number of nodes increases.
	Hierarchical: lower latency in long distance routing, lower routing overhead in large networks	The combination of proactive and reactive type of routing in a hierarchical structure is complex to implement.

Table 2 Types of Routing

Routing algorithms used for routing in *conventional systems* can be broadly classified as *static* and *adaptive* routing algorithms [66]. **Static routing algorithms**, e.g. the optimal principle, Dijkstra algorithm [66] also known as shortest path or forward search routing algorithm, perform well as long as the network status does not change. Routing information is gathered in the network before sending any data, hence this type of routing has lower latency[66]. The drawback of these types of algorithms is that once the route table is determined, it does not change in response to network changes.

Adaptive routing algorithms allow a station or a node to respond to network changes and update its routing tables accordingly. However they are slow when the size of the network grows. The Bellman Ford algorithm [66], also known as the distance vector or backward search algorithm, was initially developed for centralised systems, and later evolved into a distributed version for service points that is connected to a fixed infrastructure [66]. In this kind of system, every station learns route information only from its neighbours, and works in reverse order as listed below:

- Each node calculates the distances between itself and all its neighbours, and then stores this information in its routing table.
- Each node sends its table to all neighbouring nodes.
- When a node receives distance tables from its neighbours, it calculates the shortest routes to all other nodes and updates its own table to reflect any changes.

In contrast, WANET uses **distributed routing**, everything that centralised routing does in conventional wireless systems, e.g. gather routing information, to determines the best routes, construct routing tables etc, is executed in WANET independently at the terminals instead of at the central stations[66]. Initially each node only exchanges information with its neighbours. These neighbours then propagate the known route information to their neighbours, and gradually all remote nodes can learn about further away destinations by receiving node-by-node propagated routing information. By adding its own cost to reach a known neighbour to the cost from the neighbour to the remote destination, it can calculate the cost to reach any remote destination, and it constructs its own routing table with a record of the best route and the cost to reach these destinations. The main disadvantage of this mechanism is:

- It has low scalability, due to the fact that routing overheads mount up as the total number of nodes in the network increases.
- Slow updates on network topology changes, due to the node-by-node information spreading.
- It may trigger a deadlock called count-to-infinity (i.e. a broken link to an unreachable node may cause the rest of the nodes to gradually increase their estimates to reach the unreachable nodes using information from their neighbours)

Depending on how their routing tables are constructed, *distributed routing algorithms* proposed for WANET can be roughly classified into three fundamental categories: the *proactive* or *table driven* routing algorithms, the *reactive* or *source initiate on-demand* routing algorithms, and the *hierarchical* routing. [2, 6, 8, 9, 11, 13, 15, 29, 33, 36, 67-70]

The **proactive**, known as table-driven, distributed routing algorithms are similar to the static routing algorithms, in that they constantly maintain routing information, and routes are selected and stored even before they are needed. The advantage of proactive algorithms is that they have shorter initial route discovery delay, and select a route from the routing table without initial route discovery each time. The drawback of proactive algorithms is additional routing control traffic, which means each router wastes bandwidth to maintain routes even when it is not in use.

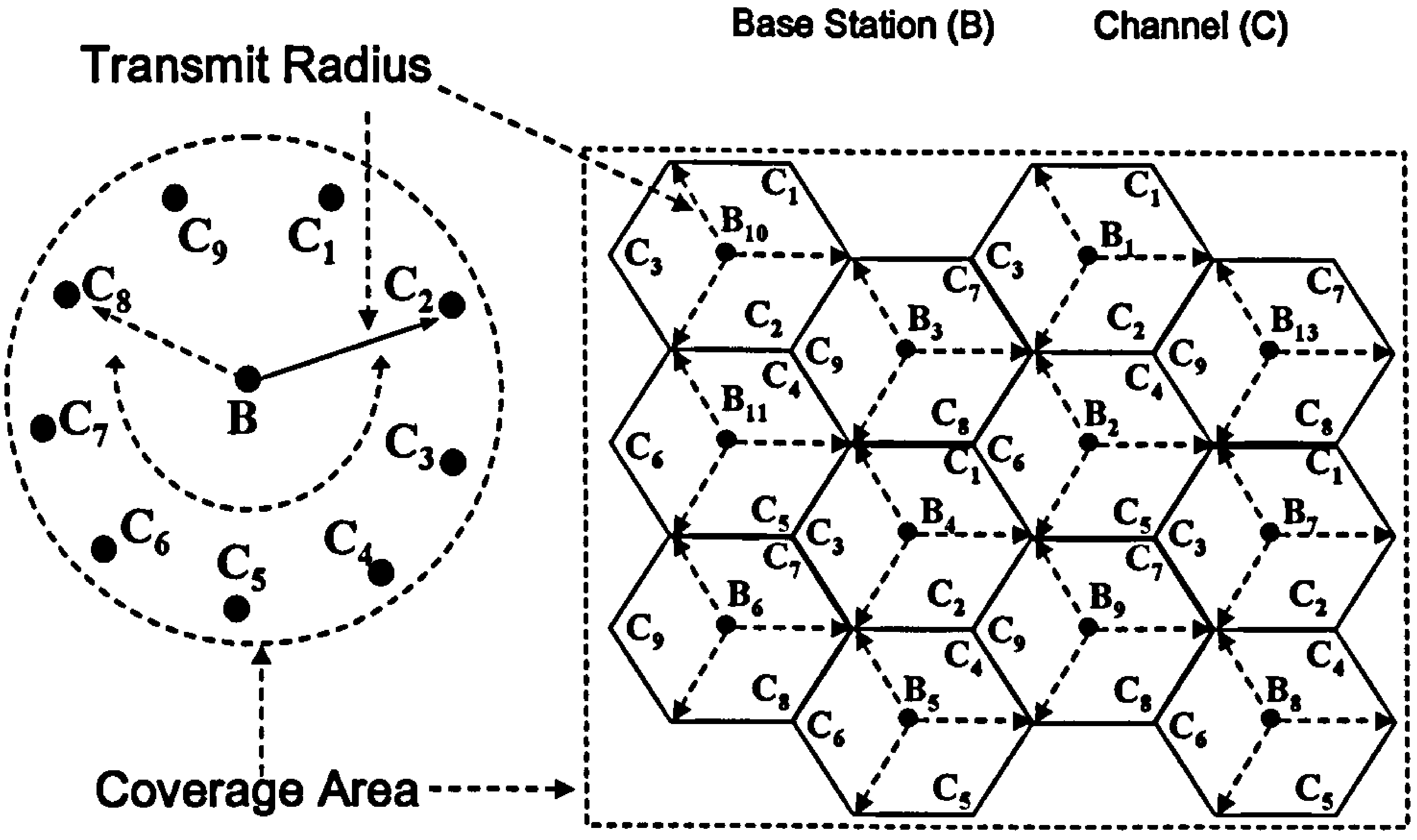
The **reactive**, also called source-initiated on-demand distributed routing algorithms, activates a route discovery procedure by the source node, and the routing tables do not maintain routes to all destination nodes all the time. The established routes are maintained until the destination becomes inaccessible via every path or the route is no longer required. The advantage of on-demand routing algorithms is that the bandwidth consumption to maintain the routing table in each node is far less than with table driven algorithms, and it is loop free. The drawback of reactive algorithms is the longer initial route discovery delay, and low scalability due to the routing overhead expended as the number of nodes in the network increases.

The **hierarchical** routing algorithms are based on the concept of a cluster-based hierarchical network structure, in which all nodes are grouped into a set of clusters. A cluster head is responsible for scheduling transmissions and resource allocation between one or more peripheral nodes within the same cluster. A cluster head then uses proactive routing to manage communication within the cluster, and uses reactive routing to communicate with nodes in the far distance. The advantage of hierarchical routing is the reduced routing overhead for long distance communication, and shorter latency for short distance communications. However, this type of routing algorithm is more complicated to implement. Chapter 3 discusses the distributed algorithms for WANET routing in more detail.

2.6 System Level

The core issue of this research is to eliminate the combination impact of interference from system level, and improve system scalability by improving **spatial reuse**, which has a similar objective for frequency reuse in a cellular structure in conventional wireless systems. Early mobile communication networks, e.g. the United States Bell mobile phone system developed in the 1970s, used a high power transmitter broadcast, with an antenna mounted on a high tower to achieve a large coverage area. This resulted in very low spectrum efficiency and user capacity of approximately twelve simultaneous calls over a thousand square miles, due to the fact that the frequency channel cannot be reused in the same area.

Instead of use a single high power transmitter, the later cellular systems (e.g. GSM) use multiple lower power transmitters, where each one only manages a portion of the total number of channels available to the entire system. Neighbouring base-stations are assigned a different set of channels, so that all available channels are assigned to a cluster of (e.g. 3, 5, 7 etc) neighbouring cells. This is achieved by systematically spacing the co-channel cells, and repeating this cluster as many times as necessary to cover a large area. In this case a cellular system can achieve high user capacity, up to thousands of users, by carefully organizing the same set of frequency channels to be reused at a safe distance.



(a) (b)

Figure 6 Wireless Networks Topology

(a) Early Mobile system (b) Cellular System

Figure 6 (a) and (b), shows the coverage area and the transmission radius differences between the early mobile systems and the later cellular systems, frequency reuse in the cellular structure with centralised control solves the scalability problem in a centralised system. However, this mechanism cannot be directly adapted in WANET, due to its highly integrated system structure. The challenge raised here is: can WANET develop a similar mechanism to provide high spatial and frequency reuse? More discussion about this issue will be presented in chapters 7 and 8.

2.7 Conclusion

This chapter has provided a literature review of the fundamentally relevant technical issues associated with the *scalability* problem for WANET capacity enhancement. The peculiar nature of WANET has influential impact on both the identified application potential and associated technological challenges. WANET is fundamentally different from previous wireless communication systems, and it has to deal with interference differently in an infrastructureless wireless transmission environment, with resource constrains. These are the sources of all WANET technological challenges.

The *infrastructureless structure* of WANET requires distributed control, meaning functions of a central location in a conventional wireless system now have to be performed by wireless terminals. The *dynamically changing link and network characteristic* demands a higher degree of inter-terminal cooperation and organization. The *limited resources*, e.g. energy and power, have to be allocated and shared fairly with higher efficiency. Most importantly, as the radiation radius of each wireless controlling unit becomes shorter, the burdens of relaying and scheduling have made the user population of a system even more difficult to scale, hence the low scalability.

Power control schemes of the physical layer, can improve energy conservation, reduce interference, increasing the number of simultaneous transmissions, and consequently improve the user capacity of a system by improving spatial reuse. *Contention control* algorithms of MAC layer, such as CSMA-CA/D, spread spectrum access (e.g. FHSS, DSSS/CDMA), and polling, are proven capable of resolving collisions in a contention based wireless environment that is shared by all wireless terminals. *Distributed routing* *algorithms*, with the capability of managing dynamic network status, can establish the robust connection that a system needs for various types of traffic.

The foundation of this research is a distributed routing algorithm, incorporating a unique *Distributed Interference Impact Probing* (DIIP) technique, collaborating with power control and dynamic topology control, to control collision and contention in a WANET environment. This combination of a set of beneficial technologies, indexed by a DIIP mechanism is aimed at enabling the system to achieve higher capacity. This mechanism will be further discussed with more detailed analysis in chapter 5.

Chapter 3 Overview of Routing Strategies

Contents

3.1 Introduction

Routing is often referred to as a two part process: one is identifying the existence of the destination nodes (called *Addressing*), used for exchanging local topology information for establishing and maintaining optimal connections between two communication points (also called *Topology Control*); the other is to construct a routing table in each router (called *Route Discovery*), which looks up a path from this routing table to direct outgoing traffic to determine the next-hop router (called *Routing*). Routing protocol is the recognized standard of algorithm support by all subsystems that share the same standard, e.g. the Routing Information Protocol (RIP) [34] and the Open Shortest Path First (OSPF) [71] routing protocol etc.

This chapter presents an overview of routing, in two major sections: the *routing environment* (relevant limitations, challenges), and a summary of representative *routing algorithms*. Each communication network can be considered as constructed by many routers, which are interconnected by links with a function of initiating or processing traffic. Each router in the following article, without a specific explanation, is referred to as a node in the network topology. Routing is a network layer responsibility as illustrated in Figure 7, in context of the OSI reference model [22].

The general goal for developing a distributed routing algorithm for WANET is: identify the wireless nodes, establish a network connection in a dynamically changing topology, and frequently update their routing table for possible changes of existing connections. The challenges for achieving such an objective include minimizing routing control overheads, reducing the route initiate latency, maximizing the network capacity, and improving the routing efficiency, in a distributed manner. In order to achieve this goal, the WANET routing algorithm design should consider various factors in the complex wireless operating environment, as well as the distributed inter-node cooperation mechanism.

Routing algorithms are characterised by the way they obtain routing information to form the routing table or their route selection criteria. In communication systems with centralised controls, routing algorithms can be classified into two categories: static/non-adaptive and dynamic/adaptive. Static algorithms make a routing decision not based on the most up-to-date network status measurements, e.g. traffic or topology, and do not reflect network changes. The simple routing algorithms such as shortest path routing, flooding and flow-based routing algorithms are in this group. Adaptive algorithms, on the contrary, make routing decisions referring to up-to-date network changes. Distance Vector Routing (DVR), link state routing algorithms, and numerous centralized routing algorithms derived from them belong to this category.

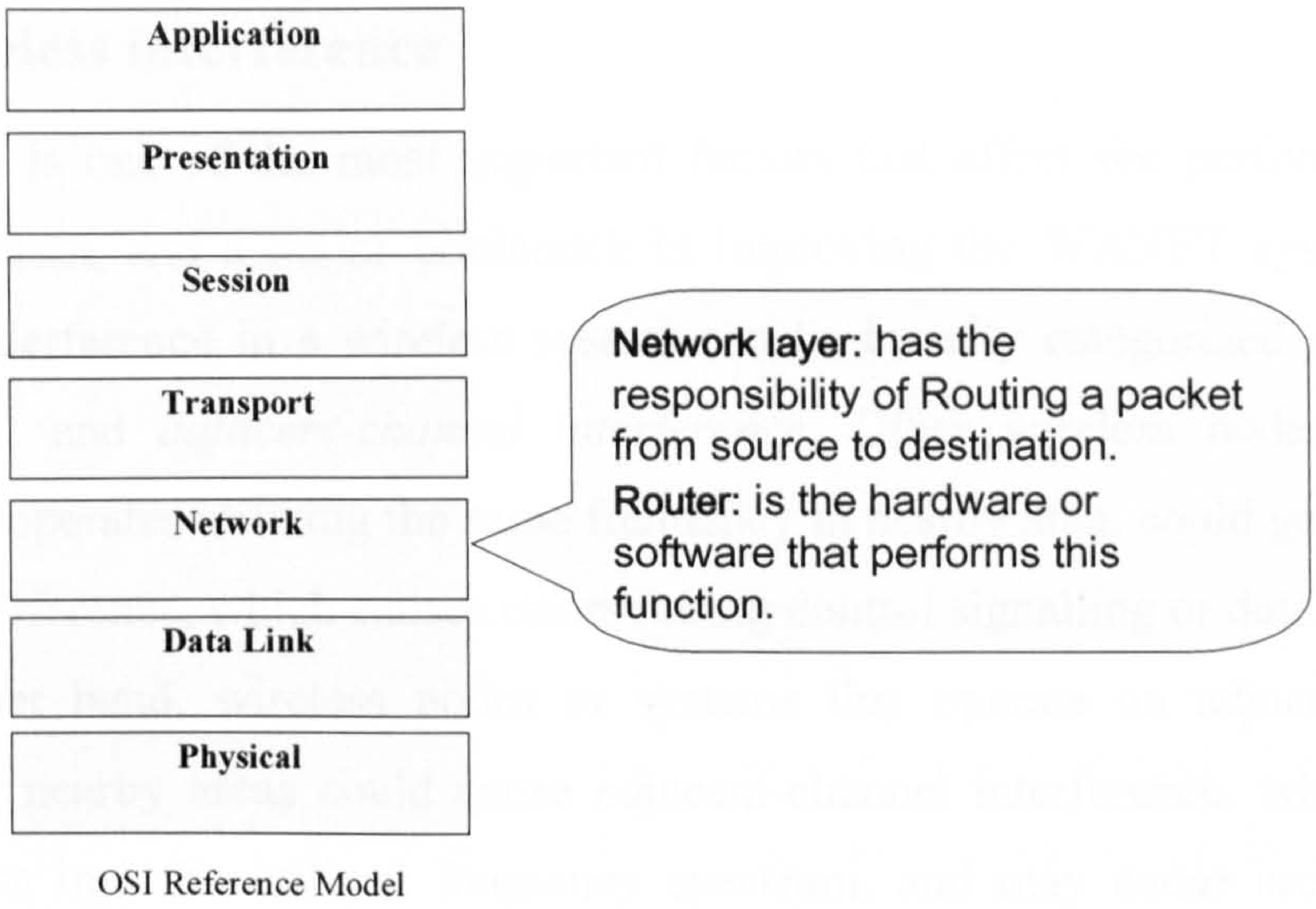


Figure 7 Network Layer in Open System Interconnection Model

Distributed routing algorithms can be summarised, depending on how their routing table is constructed, into three basic classes: proactive/table driven, reactive/source initiate on-demand, and hierarchical/hybrid.

This chapter will first introduce the routing environment, which includes key elements (e.g. radio interference, transmit power, link length, and node density etc.) that affect the routing operation. A brief summary of classical centralized (static and adaptive) and distributed (proactive, reactive, hierarchical etc) type routing algorithms is given.

3.2 Routing Environments and Strategies

WANET inherits the complex wireless communication environment characteristics of conventional wireless communication systems, therefore WANET also suffers from similar problems. The development of multi-function wireless data communication devices, and the increasing available bandwidth, make realization of WANET possible in the near future [13]. However, there are still challenges that remain unmet before WANET can be deployed for service. Before the overview of routing algorithms, it is necessary to understand the environment that a routing algorithm may be operating in. In terms of routing environment, the *wireless interference*, *variable transmitting power*, *variable link length*, and the *node density*, are the four most influential key elements in the performance of a routing algorithm.

3.2.1 Wireless interference

Interference is one of the most important factors that affect the performance of the wireless system, and a major bottleneck in improving the WANET system and user capacity. Interference in a wireless system can be broadly categorized as *co-channel interference*, and *adjacent-channel interference*. Other wireless nodes, or another system that operates utilizing the same frequency in nearby area, could generate the co-channel interference, which causes errors during control signalling or data transmission. On the other hand, wireless nodes or systems that operate on adjacent frequency channels in nearby areas could cause adjacent-channel interference, which results in energy leaks into the utilized frequency spectrum, and may cause cross-talk at the receivers.

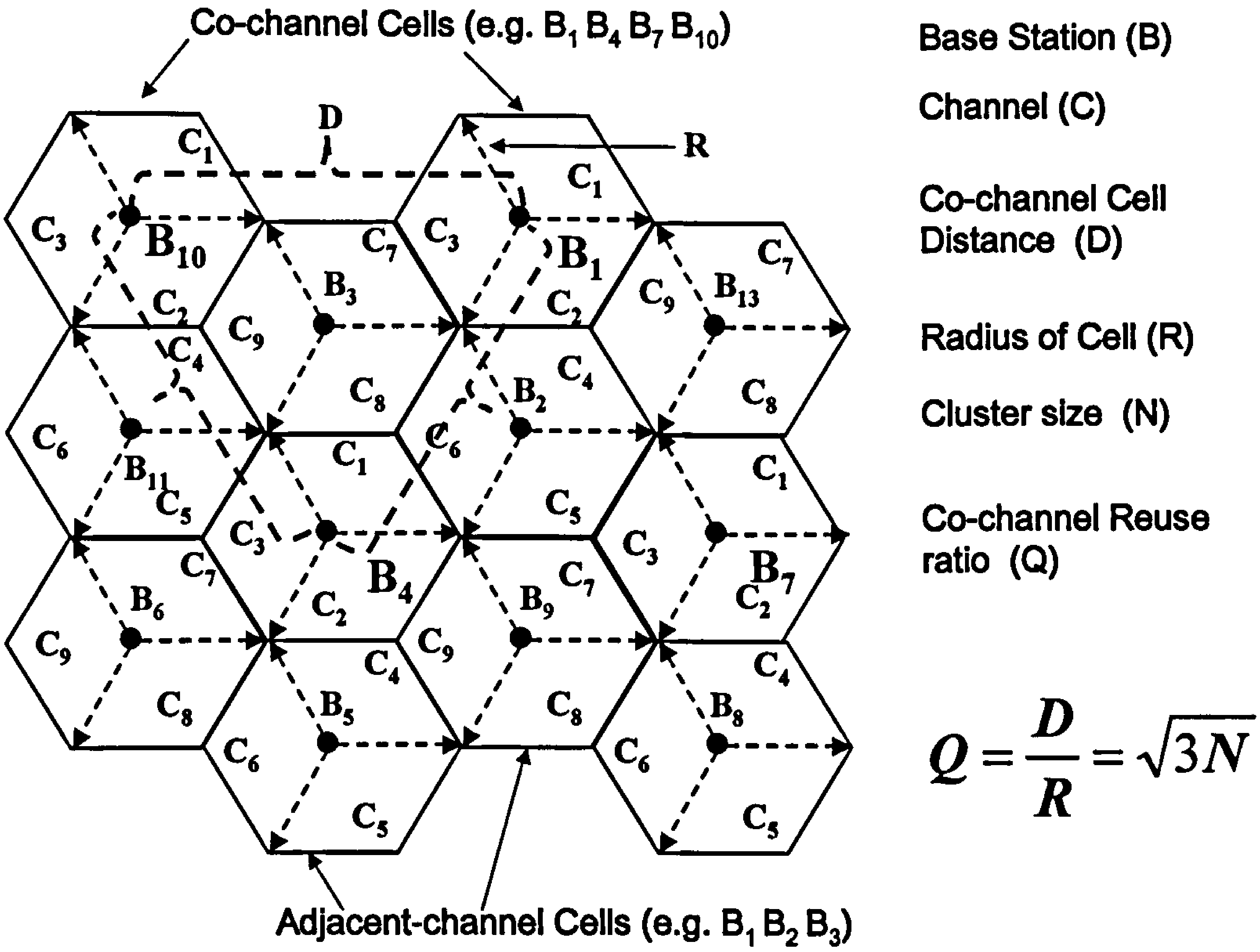


Figure 8 Frequency Reuse in a cellular structure [43]

Co-channel interference (between nearby nodes or systems) could be overcome by physical separation of co-channel terminals with a minimum distance to provide sufficient isolation. Conventional wireless systems, such as the cellular systems shown in Figure 8, manage co-channel interference by utilizing a *frequency reuse* mechanism. This mechanism carefully organizes a group of 'N' base stations, which each use part of the available frequency spectrum in each cell, and a few of these cells form a cluster. This cluster is then repeated for as many times as necessary to satisfy the user capacity, whilst making sure those co-channels cells are not deployed near each other, as shown in Figure 8.

In this example, base station B_1, B_4, B_7, and B_{10} are co-channel cells. Each cell has a coverage area with a radius 'R', and each co-channel cell is separated from any other co-channel cells with a distance of 'D'. If we assume each cell in the system is approximately the same size and uses the same transmit power level, then the co-channel reuse ratio 'Q' is the function of the radius of the cell (R) and the minimum separation distance of two nearest co-channel cells (D). In this case, in the same physical area, if the cluster size/number of cells N is small, the user capacity of the system will be lower, but each user may receive higher transmission quality. If N is large, which means the R of each cell is smaller, the user capacity will be higher. However the transmission quality may be poor due to each user having a smaller proportion of the bandwidth utilization.

By contrast, adjacent-channel interference, from nearby nodes in the same system, or other nearby systems, is far more difficult to predict and manage. Imperfect filtering mainly causes this type of interference, in which case it results in receivers that are confused and have difficulty in distinguishing weaker transmitting terminals from strong crossover adjacent-channel terminals, hence resulting in errors in decoding signals.

This is similar to the co-channel capture effect, or so-called near-far problem in wireless systems. There are different means that could minimize the adjacent-channel interference, for example: use of better filters or careful filtering mechanisms to prevent cross-over signals; use of power control to reduce interference, and hence reducing errors; and a channel assignment strategy to keep frequency separation between each channel in a given cell as large as possible [43].

Both the co-channel and the adjacent-channel interference from wireless terminals could cause significant impact on the system and user capacity in WANET, due to its distributed nature. Therefore WANET needs a high degree of cooperation and interoperation both within the system and between other systems. Within the system a frequency-reuse mechanism, similar to a cellular system, is essential to solve the scalability problem in WANET, and deliver high system and user capacity, in terms of proving adequate channel capacity serving mass number of users. Between co-existing adjacent and co-channel systems, a highly cooperative protocol will be the foundation of future wide deployment of WANET in any noisy environment, such as in urban areas.

3.2.2 Transmit Power

In WANET, the transmit power of each wireless terminal could be controlled, so that it can adapt to the network status changes. However, the variation of transmit power causes a series of impacts in a wireless environment, such as variation of *link length*, changing network connectivity, and formation of a dynamic network topology. All these variations, each pose a challenge for routing in WANET [3, 16, 18, 20, 21, 45, 72].

The *Variable Transmitting Power* (VTP) allows terminals to use minimum transmit power to achieve adequate Signal-to-Interference and Noise Ratio (SINR) at the receiver side, hence accomplishing the purpose of energy conservation. By adjusting the transmit power level, the interfering range of each transmitter becomes adjustable thereby reaching only the desired receiver, avoiding interference on irrelevant neighbouring nodes or the near-far problem, allowing more simultaneous transmissions to take place, therefore achieving higher spatial reuse. The VTP mechanism is also useful in a high node density environment. In some cases, the VTP is used simply because of insufficient remaining battery power at a wireless terminal, so instead reaching the nearest neighbour to carry out a minimum distance communication becomes a priority.

As a result of variable transmit power, the wireless link-lengths will change accordingly, in which case it will pose a chain reaction concerning routing overhead increases in WANET. Firstly, variable link length triggers a variation in the *number of* *relays* made by intermediate nodes, i.e. the *number of hops* between a pair of communicating nodes. In this sense, the increase and reduction in the number of hops will affect any routing algorithm that uses hop-count as the routing decision criteria.

Consequently if more intermediate nodes are involved in a routing operation, the routing control information exchanges between them will increase proportionally, hence the routing overhead may increase accordingly. Secondly, the variation in link-length will cause the connection or disconnection of wireless terminals, which depending on the situation, hence the network topology will change, which means the routing update as a result of topology changes will also occur more often, and the consequence of this is an increased routing overhead.

Both the increased number of relays and frequent topology updates, caused by variable link-length due to transmit power changes, results in routing overheads increasing proportionally, even dominating the network traffic. Hence control of the transmit power is no doubt an unavoidable issue in WANET routing that should be addressed of with serious consideration.

3.2.3 Link Length

Variable transmit power causes variation in wireless link length and has significant impact on routing, in terms of overheads and system capacity, which will affect the sustainable user population and the transmission quality [16].

In terms of system connectivity, the long link length enables the transmitting nodes to reach destination with fewer hops or relays, and consequently less relaying burden and delays.

For system throughput, the increased link length, means high transmit power, and stronger signal energy, however, it also implies high interference and noise, hence the overall throughput and SINR will decrease as a result.

In terms of routing, longer link lengths bring benefits such as a smaller number of relay overheads. The adverse short link length may suffer a high proportion of relay overheads. In reference to the above reasons, one can conclude that an optimal link length is essential both for enhancing the system throughput and for routing in WANET.

3.2.4 Node Density

In a network, the same routing operation with different node densities could produce completely different result [73]. In WANET, a high node density means higher network connectivity, as well as more interference, more interruption, and more contentions between nodes, hence higher routing and control overheads, and a higher chance of collisions. Strong connectivity is beneficial for maintaining a network connection for routing but the high interference further complicates the wireless environment. Frequent contention for transmission triggers an increasing number of collisions.

The potentially increasing hops and relaying, due to the increased number of intermediate nodes on a multi-hop path, generate more routing maintenance and relay overheads, which could increase proportionally as the number of node increases asymptotically. Eventually if the node density increases to a dense enough situation, the routing control traffic will simply dominate the network traffic, in which the network will transmit no data traffic, but only routing overheads. On the contrary, the low node density implies weaker network connectivity, lower interference, less contention, fewer collisions, and fewer relay overheads. However, the increasing demand on WANET user capacity means the node density in such a network will in most cases, asymptotically increase.

It has been suggested that variable transmit power could deliver the optimal link length, hence system throughput [73]. Further more, an optimal number of surrounding nodes, which form a strong enough connectivity, will enable a transmitting node to achieve optimal throughput.

However in a random network where wireless nodes are randomly distributed, if one node has determined its optimal number of neighbouring nodes, can other nodes be satisfied concurrently? There exists a trade-off between optimal node density and optimal transmit power, which also influences the number of hops for relaying in the network. Any routing algorithm design should ignore such an issue, since the important interrelationship between optimal transmit power, optimal link length, optimal neighbouring node density, and optimal number of hops for relaying, can significantly affect the performance of a routing algorithm, and furthermore the performance of the wireless network.

3.3 Routing Algorithms

Routing algorithms for wireless communication systems have evolved from the *centralized* era to *distributed* era. Depending on where the routing process takes place within a system, existing routing algorithms can be broadly summarised into two categories: centralised routing, and distributed routing.

Centralized routing algorithms operate in a central location, such as a base station in the cellular system, where all the function and burdens of routing are carried out by this central location in a centralized manner [1, 2, 4, 9, 12, 14, 22]. Most conventional communication systems use this type of routing. Depending on the adaptability to network status changes, centralised routing algorithms can be classified as static or adaptive algorithms. In the centralized stage of routing algorithm development, *static* routing algorithms are designed to determine and maintain connection in relatively stable networks, where status changes are rare. The *adaptive* type of routing algorithms are designed for networks which have variable connection conditions, but based on centralized controls, such as infrastructure based wireless systems [42, 43].

Distributed routing algorithms, in contrast, operate at terminals, such as wireless nodes in WANETs, and carry out the burden of routing and function as a router, in a distributed manner [1, 12, 14]. Communication networks, such as WANET, tend to develop wireless systems that operate without the support of fixed infrastructure, whilst it remain inter-connectable to existing infrastructured cable or wireless systems. Depending on the method used for constructing their routing table, this type of routing algorithm can be classified as proactive, reactive, hierarchical etc.

Proactive routing algorithms evolved from earlier centralized algorithms, for distributed operation with short initial route discovery delay. However, this type of algorithm has very high route control overheads, is due to the proactive maintenance of routes that are not in use. The *reactive* type of routing algorithms can reduce these routing overheads, since they only initiate routes when needed, and discard them when not in use, but these algorithms can take too long to converge. *Hierarchical* routing algorithms combine the proactive and reactive algorithms, and this type of algorithms have less routing overhead than proactive algorithms, and furthermore they converge faster than reactive algorithms. Miscellaneous distributed routing algorithms are derived from these three basic routing categories.

3.3.1 Static Routing Algorithms

The static/non-adaptive routing algorithms make routing decisions and compute routes in advance, which means the route is chosen and stored in each router long before the communication starts. The path leading to the destination node is pre-calculated and depends on static network topology and pre-estimated traffic information. However this pre-stored routing information does not reflect any network status changes that take place over any period of time.

3.3.1.1 Optimality Principle and Shortest Path Routing

The **optimality principle** is one of the basic static routing algorithms that select the optimal path according to a network topology map. The map is plotted when the network has just started. Each node draws an optimal route map that is rooted from itself, and records all optimal routes leading to every known destination in the network. This tree shaped map is called a **sink tree**, as shown in Figure 9.

Figure 9 (a) illustrates an overall network topology map, which contains nodes, indicating routers, and arcs, indicating links. Figure 9 (b) is the sink tree plotted by the node A, in which case from node A to node E there are two alternative paths, the A-H-E and the A-H-F-E, whereas the optimal path in the map according to the optimality principle is the shorter route A-H-E. Figure 9 (c) and (d) is the sink tree rooted from the source nodes B and C respectively, and shows paths leading to all destinations in the network.

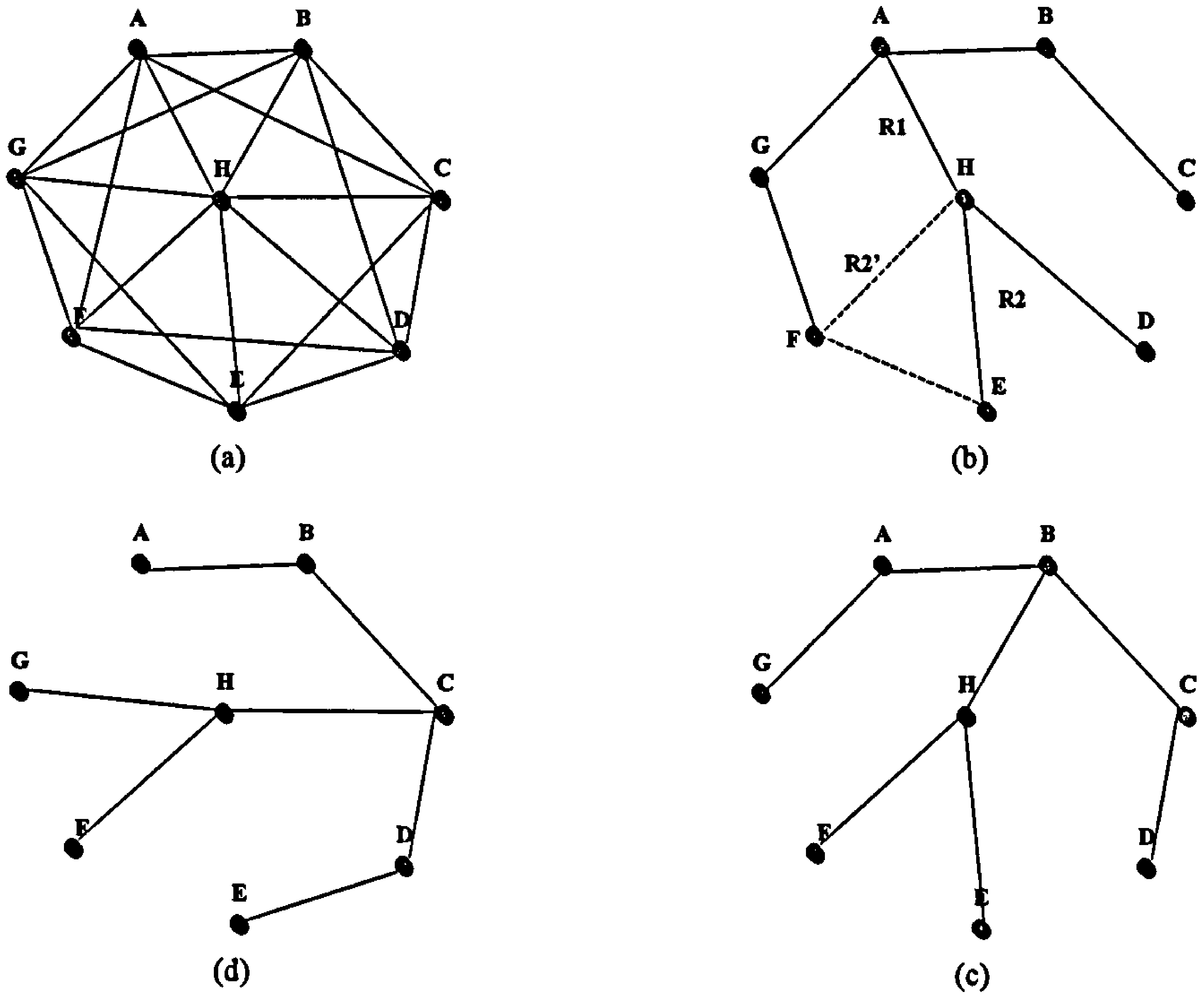


Figure 9 Optimal Principle

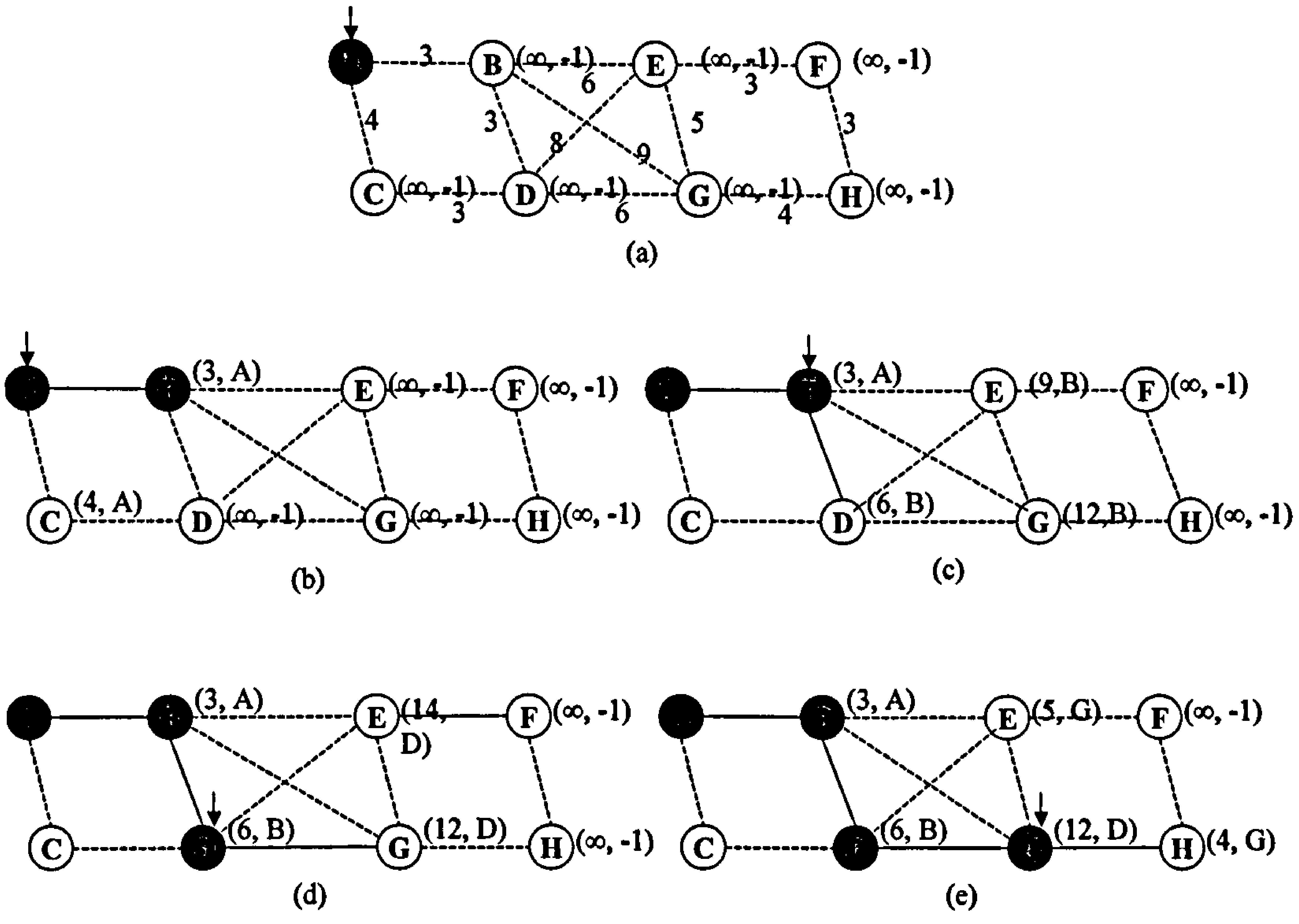
(a) Subnet Map (b) sink tree of node A (c) Sink tree of B (d) sink tree of C

Shortest Path Routing (SPR) is the most fundamental routing algorithm jointly used in many different forms because of its simplicity [1, 13, 42, 74-76]. The idea of the SPR algorithm is to construct a graph of a communication network, where each link is labelled with a measure in the graph. In this case a route weight can be calculated as the function of these measures, such as distance in the number of hops, mean queue size, bandwidth, average traffic load, transmission delay, communication cost, mean queue length and other factors.

The **Dijkstra** algorithm (Dijkstra 1959) is one of the numerical algorithms used for computing the shortest path in SPR [76]. The measuring of a shortest path can be one criterion or a combination of different criteria, whilst the selection of a route in some cases could be the fastest path rather then shortest distance. After all the links have been measured and labelled, the graph of the network is ready for a particular routing operation. Figure 10 illustrates the first four steps of finding a route from node A to H using SPR.

Figure 10 (a) shows the whole map of a network of eight nodes, each interconnected with neighbouring nodes by links labelled with an initial weight. Figure 10 (b) illustrates the first step inspection of all adjacent nodes to the working node A, where there are two alternative next-hop nodes, B or C, and because B has a shorter link, it will be marked as the working node in the next hop.

Figure 10 (c) shows the second step, node B repeats the previous inspection and marks D as the next working node. Figure 10 (d) and (e) are the third and fourth step, where each working node repeats this router discovery operation until the destination is reached. Each node along the shortest path is labelled with a distance to the source, A, and the last hop node that it passes through.



(a)

(b)

(c)

(d)

(e)

Figure 10 Shortest Path Routing

(The first four steps of route discovery [22])

The optimality principal and SPR introduces the basic method of selecting the shortest path, and the rest of section 3.3.1 introduces another basic routing algorithm called flow-based routing, which reflects network status changes.

3.3.1.3 Flow-based Routing [22]

Routing algorithms introduced so far only take network topology into account, ignoring the traffic loads, whereas Flow-Based Routing (FBR) considers traffic flow when making a route decision, since traffic load variations may also affect routing decisions [28]. Imagine a chosen shortest path has heavy traffic on part of its links, in which case it may be better to choose a longer path with less traffic, hence shorter delay. FBR uses mean packet delay, which could be affected by traffic loads, on each links to measure the shortest path, in other words, the shortest path is the series of link with the shortest time delay caused by traffic loads, rather than the shortest distance.

It is possible to analyse the stable and consistent traffic flows mathematically in a network, and compute the average packet delay. Because the average packet delay reflects the traffic load on a particular link, a route selection decision could be made based on measures such as average link delay. The mean packet delay can be calculated as a fraction of the traffic load on a link. Assuming the traffic flow is “λ” in packets/second, and the link capacity is “C” in bits/second, $1/\mu$ is mean packet size in bits. Then the link capacity in packets/second is “$C/(1/\mu) = \mu C$”. The mean packet delay on a link represented by “T” can then be calculated using following equation: “$T= 1/(\mu C - \lambda)$” (e.g. if $1/\mu = 800$ msec, $C = 20,000$ bits/sec and $\lambda = 14$ packets/second, the mean packet delay T is 91 msec).

The drawback of FBR is that routing information recorded in the routing table will not change, due to its static routing nature, in which case later routing decisions cannot reflect traffic loads and network topology changes. The static routing algorithms use the estimated routing information to make a routing decision, but this cannot satisfy the growing demand on network size, complexity, traffic load, and topology changes. The later developed dynamic/adaptive routing algorithms make a routing decision referring to calculations using the most up-to-date routing information such as network topology changes or traffic loads changes [22].

3.3.2 Dynamic Routing Algorithms

The evolution of modern networks poses challenges such as fast changing network status, hence using static routing algorithms becomes less attractive, and routing algorithms that select a path that reflects network changes dynamically become more preferable. A *dynamic/adaptive* routing algorithm is defined in contrast to the static/non-adaptive algorithms that were reviewed in the previous section. Routers use dynamic routing algorithms to execute routing operations by communicating with their neighbours periodically.

The basic difference between dynamic and static routing algorithms is: static algorithms make routing decisions based on one-off routing information obtained at the beginning of the network operation; dynamic routing algorithms make decisions using information periodically exchanged between nodes. In general dynamic routing algorithms for centralised systems can be classified into two categories: Distance Vector Routing (DVR), and Link State Routing (LSR). The way of routing information maintained in the routers provides a distinguishing feature between these two types of algorithms. DVR algorithms maintain all routing information in a routing table, which indicates the cost to each destination and the next outgoing link towards it. The Routing Information Protocol (RIP) and its version 2 (RIPv2) belong to this category [34].

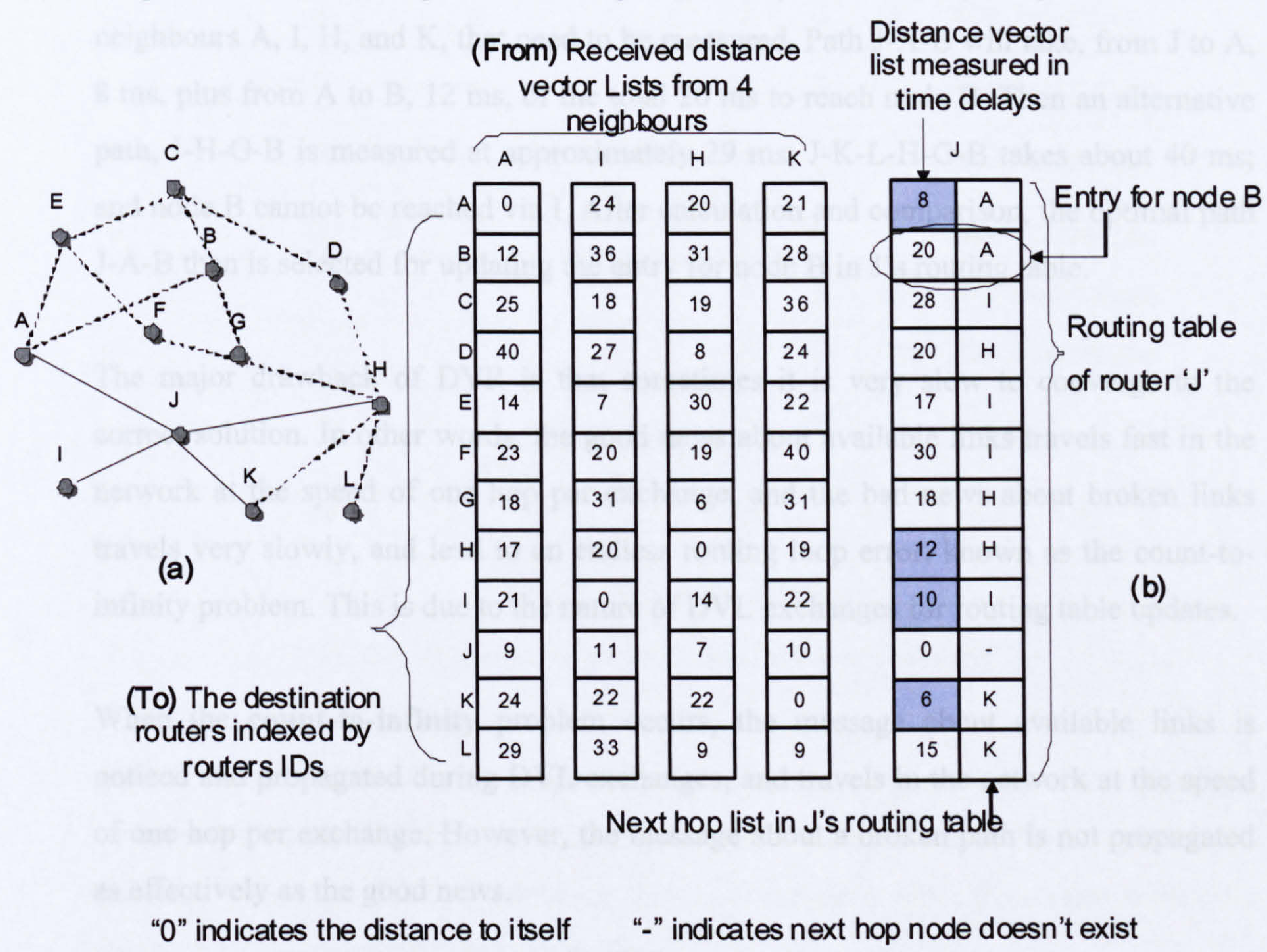
LSR algorithms maintain a buffer, which stores Link State Packets (LSP) that carry routing information of the entire network topology, whereas its routing table only stores the calculated shortest path for each known destination. The Open Shortest Path First (OSPF) [71] and Intermediate System-Intermediate System (IS-IS) [27, 28] routing protocols belong to this category. The following section describes only the common features of DVR and LSR, instead of details of each specific routing protocol.

3.3.2.1 Distance vector routing

The classic DVR uses the Distributed Bellman-Ford (DBF) algorithm, developed long before the Internet existed [22]. The modifications of DVR were used in Advanced Research Project Agency Network (ARPANET) for the U.S. Department of Defence, and were also applied to the later formed Internet under the name of RIP [34].

Each node in the network topology map represents a router, and it maintains a regularly updating routing table indexed by their IDs. The routing table has one entry for each node, containing two parts of the routing information of that router. The first part is the distance vector (e.g. number of hops; average delay time measured from a time stamped

ECHO packet; queue length on its outgoing link; or any other metrics), which records the route weight to reach a destination node; the second part is the preferred next-hop node leading towards that destination. Each node periodically updates its routing table by exchanging Distance Vector Lists (DVL) with its neighbours (e.g. once every ‘x’ ms), and updates its own routing table referring to those received DVLs [1, 77].



	A	I	H	K	J	
A	0	24	20	21	8	A
B	12	36	31	28	20	A
C	25	18	19	36	28	I
D	40	27	8	24	20	H
E	14	7	30	22	17	I
F	23	20	19	40	30	I
G	18	31	6	31	18	H
H	17	20	0	19	12	H
I	21	0	14	22	10	I
J	9	11	7	10	0	-
K	24	22	22	0	6	K
L	29	33	9	9	15	K

Figure 11 Routing table update process

(a) The subnet map (b) Routing table of node J and DVL from neighbours [1]

Figure 11 illustrates a network topology map, a routing table in node J, and four DVLs sent from J’s neighbours. Figure 11 (a) is the topology of the network, and node J is connected with four neighbours (A, I, H, K) by communication links. Figure 11 (b) shows the two-columned routing table of J, which records the distance measured by delay time, and a preferred next-hop node list for forwarding packets; and 4 single column DVLs sent from J’s neighbours for exchanging knowledge about further away destinations. Eventually J’s routing table will have the distance measurements to all known destinations in the network. On the other hand, J will send its own DVL to its neighbours.

For example, assuming that J needs to calculate a route to reach B, which is a node not directly connected to J, and there are four optional paths from J, via its four immediate neighbours A, I, H, and K, that need to be measured. Path J-A-B will take, from J to A, 8 ms, plus from A to B, 12 ms, of the total 20 ms to reach node B. Then an alternative path, J-H-G-B is measured at approximately 29 ms; J-K-L-H-G-B takes about 40 ms; and node B cannot be reached via I. After calculation and comparison, the optimal path J-A-B then is selected for updating the entry for node B in J's routing table.

The major drawback of DVR is that sometimes it is very slow to converge to the correct solution. In other words, the good news about available links travels fast in the network at the speed of one hop per exchange, and the bad news about broken links travels very slowly, and lead to an endless routing loop error, known as the count-to-infinity problem. This is due to the nature of DVL exchanges for routing table updates.

When the **count-to-infinity** problem occurs, the message about available links is noticed and propagated during DVL exchanges, and travels in the network at the speed of one hop per exchange. However, the message about a broken path is not propagated as effectively as the good news.

Figure 12 shows the count-to-infinity occurring process in a chain network, which consists of five routers using DVR [1, 22]. Figure 12 (a) shows the exchange process of the distance vector measured in number of hops, where all links are normally connected, and the message of the available 1-hop path via B to A propagates at the speed of 1 hop per exchange. B's neighbour C will add the length of B-C to A-B, so that C learns the route via B to A in 2 hops. After a finite time of DVL exchanges all nodes will learn the distance to A [22].

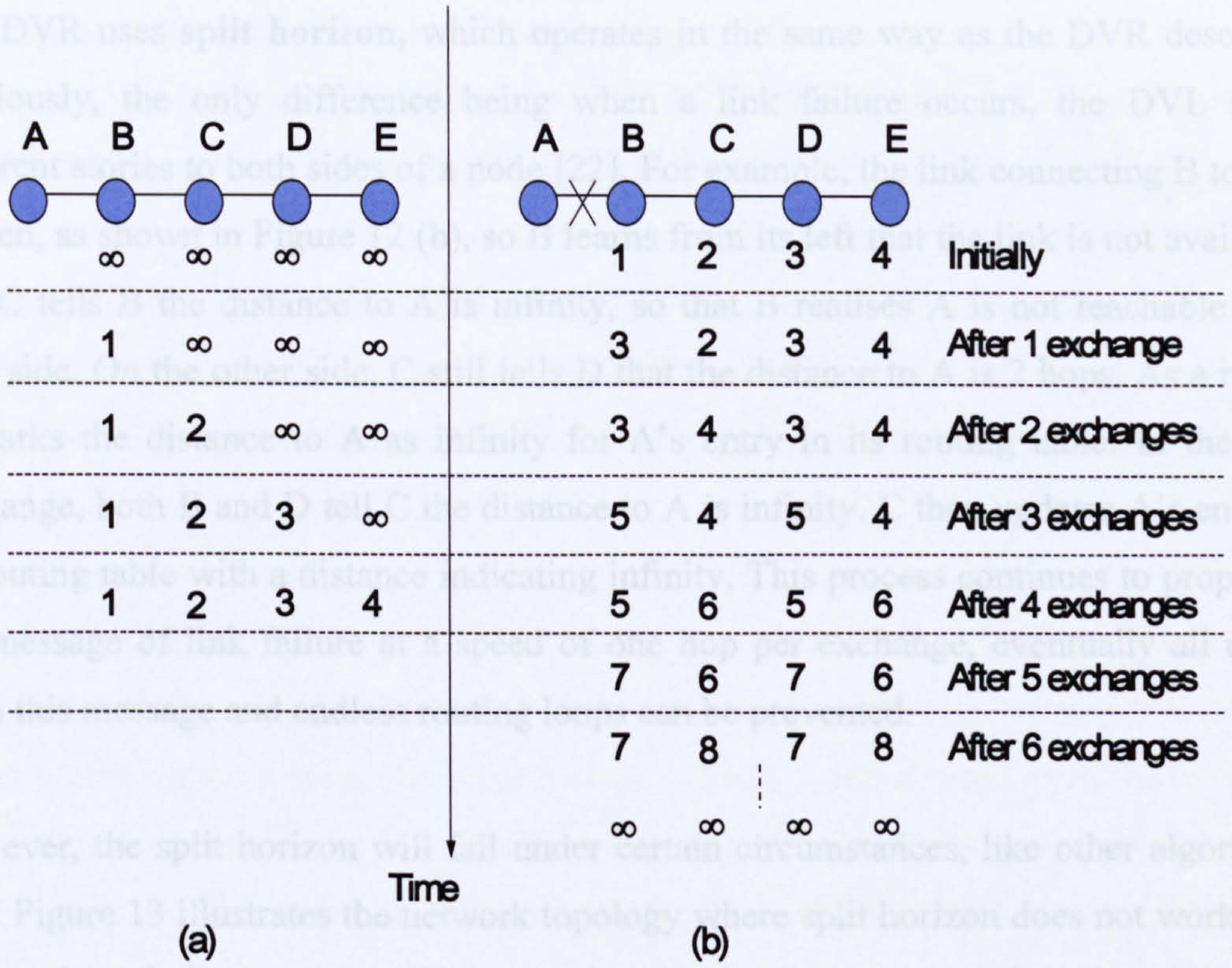


Figure 12 DVR vector exchanges

[22] (a) Normal vector exchanges (b) Count-to-infinity

Figure 12 (b) demonstrates the routing error that occurs when a broken link triggers the *count-to-infinity* loop [22]. Assuming the link between B and A has broken, the DVL exchange causes confusion for nodes in opposite directions, since the routing table only records a next hop forwarding information.

For example, B learns that the link to reach A via itself is not available, but C can reach A in 2 hops, so after 1 exchange, B updates the entry for A in its routing table for the path B to A via a route weight/cost of a total of 3 hops (B to C in 1 hop plus C to A in 2 hops). Then C later learns the same error message from B, and updates the entry for A with a cost of 4 hops. Eventually all the nodes on the other side of the broken link will be locked in a endless routing loop, and the route weight for reaching A from themselves will be increased up to infinity. This routing error will continue to loop amongst these five routers, and if the path between A and B does not come back on, the routing process will never converge to a correct answer. This problem can be solved by a method called split horizon.

The DVR uses **split horizon,** which operates in the same way as the DVR described previously, the only difference being when a link failure occurs, the DVL sends different stories to both sides of a node [22]. For example, the link connecting B to A is broken, as shown in Figure 12 (b), so B learns from its left that the link is not available, and C tells B the distance to A is infinity, so that B realises A is not reachable from both side. On the other side, C still tells D that the distance to A is 2 hops. As a result, B marks the distance to A as infinity for A's entry in its routing table. In the next exchange, both B and D tell C the distance to A is infinity. C then updates A's entry in its routing table with a distance indicating infinity. This process continues to propagate the message of link failure at a speed of one hop per exchange, eventually all nodes learn this message and endless routing loops can be prevented.

However, the split horizon will fail under certain circumstances, like other algorithms [22]. Figure 13 illustrates the network topology where split horizon does not work very well, and confusion reoccurs when more than one node operates on the other side of the broken link. For instance, router B detects the direct link to A is down, and both router C and D tell B that the distance to reach A is infinity, so that B marks the distance to A as infinity. However, D tells C it can reach A in 3 hops, so C marks the distance to reach A via D is 3 hops in distance. The following DVL exchange will continue this confusion, and another routing loop starts, eventually causing the count-to-infinity routing error.

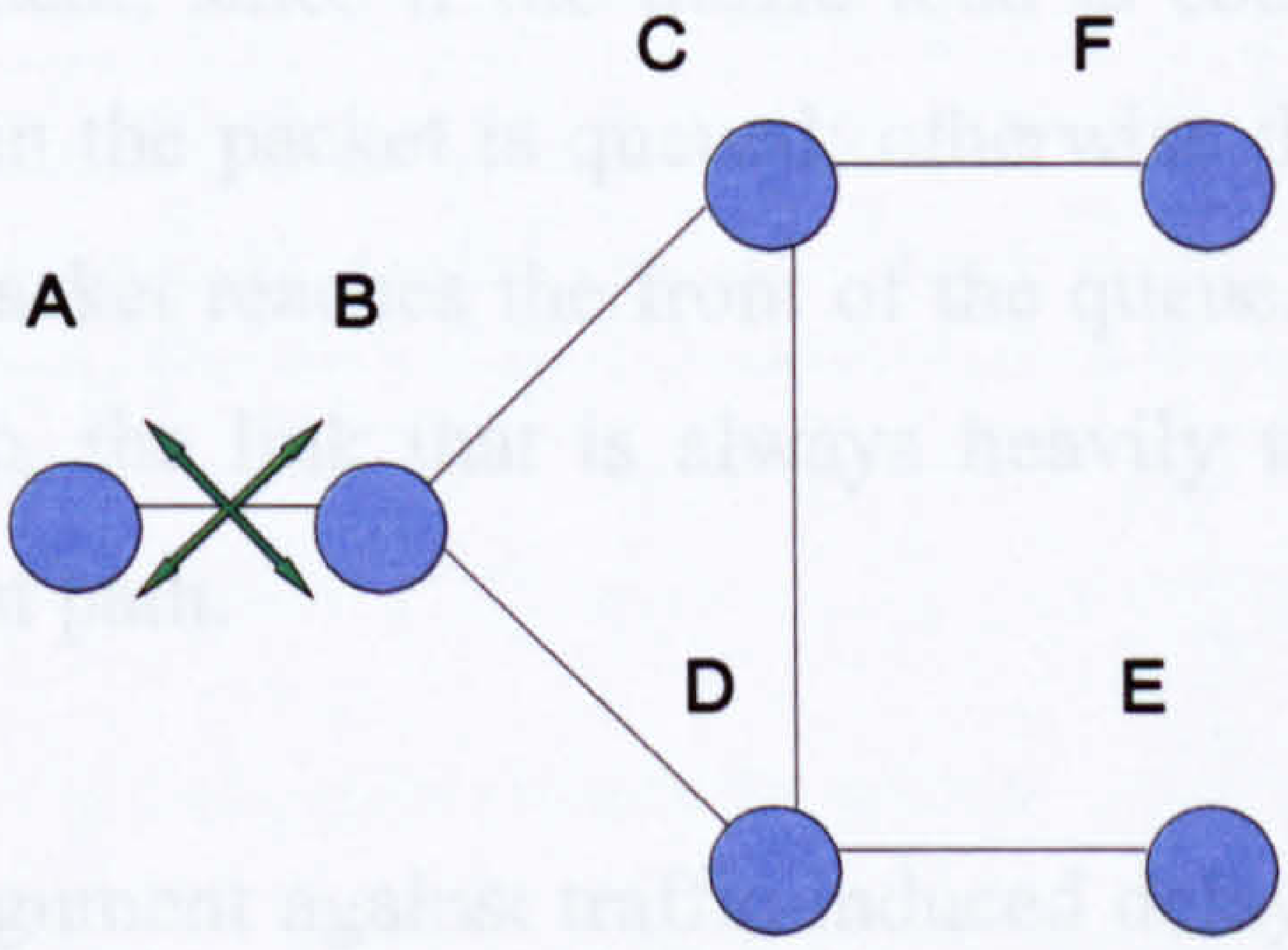


Figure 13 Split horizon hack

The DVR algorithm is efficient and easy to implement for networks that do not have many changes of topology and transmission delay. It was used in ARPANET until 1979 when it was replaced by the Link State Routing (LSR) algorithm. There are two major drawbacks which caused dismissal of the DVR. Firstly, it takes too long to converge even with improved variants like split horizon. Secondly, DVR uses the queue length as the primary metric, which does not reflect transmission speeds or bandwidth changes between different networks, whilst the transmission speed has changed greatly in the last few decades (e.g. PSTN was 56k bps; ISDN was 128k bps, DSL can transmit up to 500k bps, ATM and Broadband ISDN networks can transfer up to 150-600M bps).

3.3.2.2 Link state routing

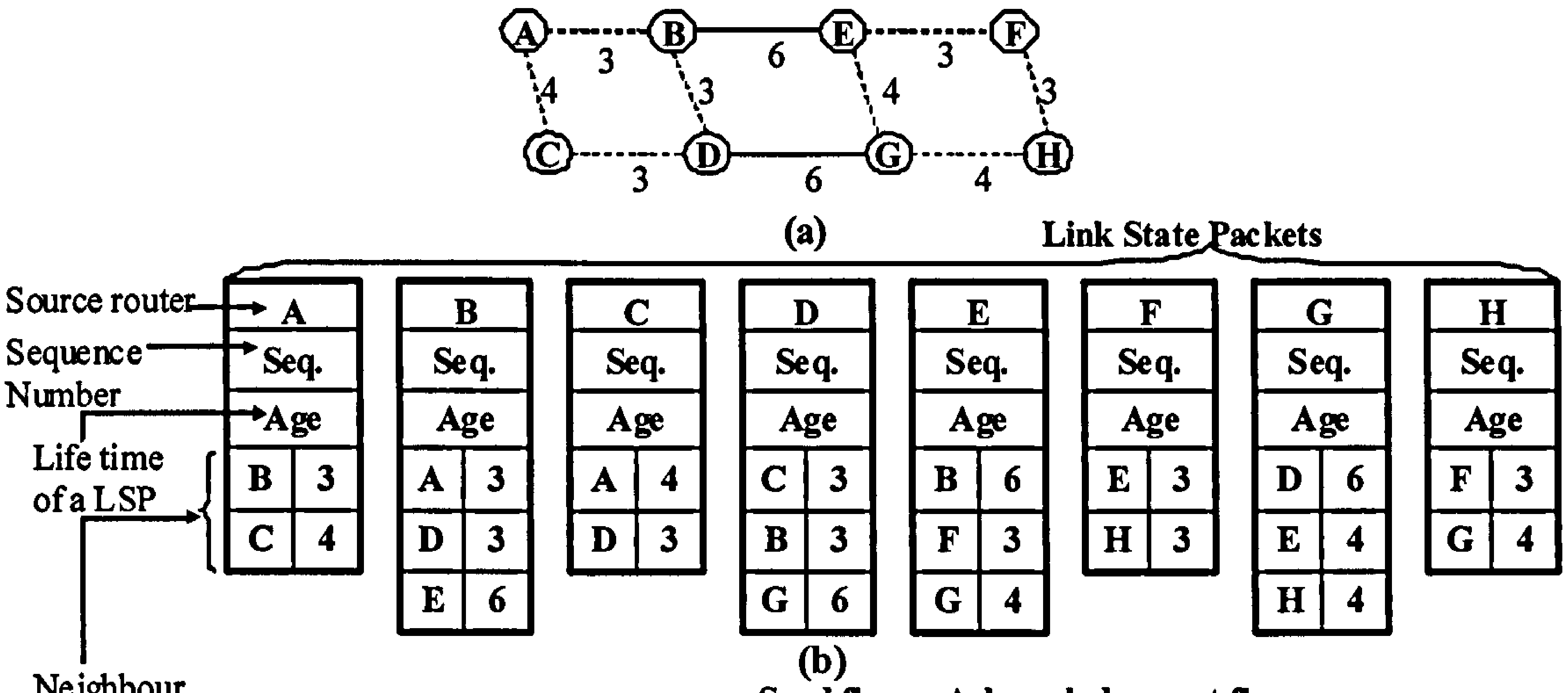
The classic link state routing (LSR) algorithm operates in two stages: route discovery and route maintenance. The first stage obtains routing information for later construction of the global network map; the second stage calculates the shortest path using certain types of measure (e.g. delay time, traffic loads etc), and updates the routing table with a route weight cost for each destination node [22]. Two stages of LSR can be summarised in five steps of operation.

The **first step** allows each node in the network to learn information about all its neighbours. In this case, each node learns which is its neighbour node and what their address identity (ID) is.

The **second step** is to measure the cost for each hop to reach all immediate neighbouring nodes. There is an argument about whether the traffic loads should be included in the measurement, since if the traffic load is counted, the round-trip time should start counting when the packet is queued; otherwise the round-trip time should start counting when the packet reaches the front of the queue. In this case, if two links have the same bandwidth, the link that is always heavily traffic loaded will not be chosen as part of a shortest path.

On the other hand, the argument against traffic-induced delays is that this method will trigger a *routing oscillation* problem. Figure 14 (a) illustrates such circumstances, where two links (BE and DG) connect 2 nodes on the left and the other 2 nodes on the right, where BE is loaded with heavy traffic, and link DG has the same bandwidth but a lighter traffic load, then DG will become the preferable route, and all subsequent traffic will congregate on this link. Soon DG becomes overloaded, and BE become vacant, so all the traffic will move back to BE. This process repeats itself and triggers an oscillating routing loop. In this sense, it is better to just measure the time delay instead of taking into account the traffic loads or bandwidth on the link.

The **third step** is to construct a Link State Packet (LSP), with a source node ID, a sequence number, an age field and a link cost table in it, so that it can record all routing information previously learnt, as illustrated in Figure 14 (b). LSPs are constructed either in a fixed period of time, or when there is a significant change in the network (e.g. a link or router goes down or comes back up again), and are propagated in the network using flooding.



Source router	Seq. No.	Age	Send flags			Acknowledgement flags			Data
			A	D	E	A	D	E	
A	20	60	0	1	1	1	0	0	
C	30	60	0	1	1	1	0	0	
D	31	59	1	0	1	0	1	0	
E	15	60	1	1	0	0	0	1	
F	21	59	1	1	0	0	0	1	
G	15	60	1	0	0	0	1	1	
H	11	59	1	0	0	0	1	0	

(c)

Figure 14 Link state routing algorithm

(a) 8 nodes network (b) link state packets (c) Packet buffer for router B

The **fourth step** is to propagate the LSP in the network to distribute routing information. This includes receiving, processing, and copying LSPs from other nodes, and then constructing a new LSP and sending it out to other nodes. The intermediate nodes receiving an LSP will store it into a local packet buffer for comparison. The intermediate routers will only overwrite the old record with a LSP if it has a newer sequence number and fresher information, and acknowledges or forwards it to its neighbours. Figure 14 (c) illustrates an example of a packet buffer, which holds information about source node ID or address, the LSP sequence number, the age or lifetime of LSP, and the matrix of "send" and "acknowledgement" flags that indicates whether a newly arrived LSP should be forwarded on to, or acknowledge neighbouring nodes. Every router stores such a packet buffer, so that all nodes in the subnet can construct a graph of the entire network. This means every node knows its neighbouring nodes and the cost to all possible destinations in the network.

The **fifth step** is to calculate the shortest path, using Dijkstra's algorithm, according to the information held in the local packet buffer, and then update the local routing table. Each node constructs a routing table, where each known source node is referenced by a routing entry which records the cost to reach a source node and the preferred next hop neighbour node. Yet LSR has a major drawback, which is the large amount of routing overheads due to the propagation and exchange of LSP, as well as the routing table that grows proportionally with network size. The next section will introduce three different types of routing, i.e. broadcast, multicast or unicast, each of which can send messages to different destinations simultaneously.

3.3.3 Broadcast, Multicast and Unicast routing

Transmitting packets from source to destination can be achieved in many different ways, such as broadcast packets to all destinations simultaneously, or multicast packets to a specific group of destinations, or unicast to a particular destination, depending on the type of application and the number of destinations. Broadcasting or multicasting is for point-to-multipoint packet transmissions, and unicasting is used for point-to-point packet transmissions.

3.3.3.1 Broadcast routing

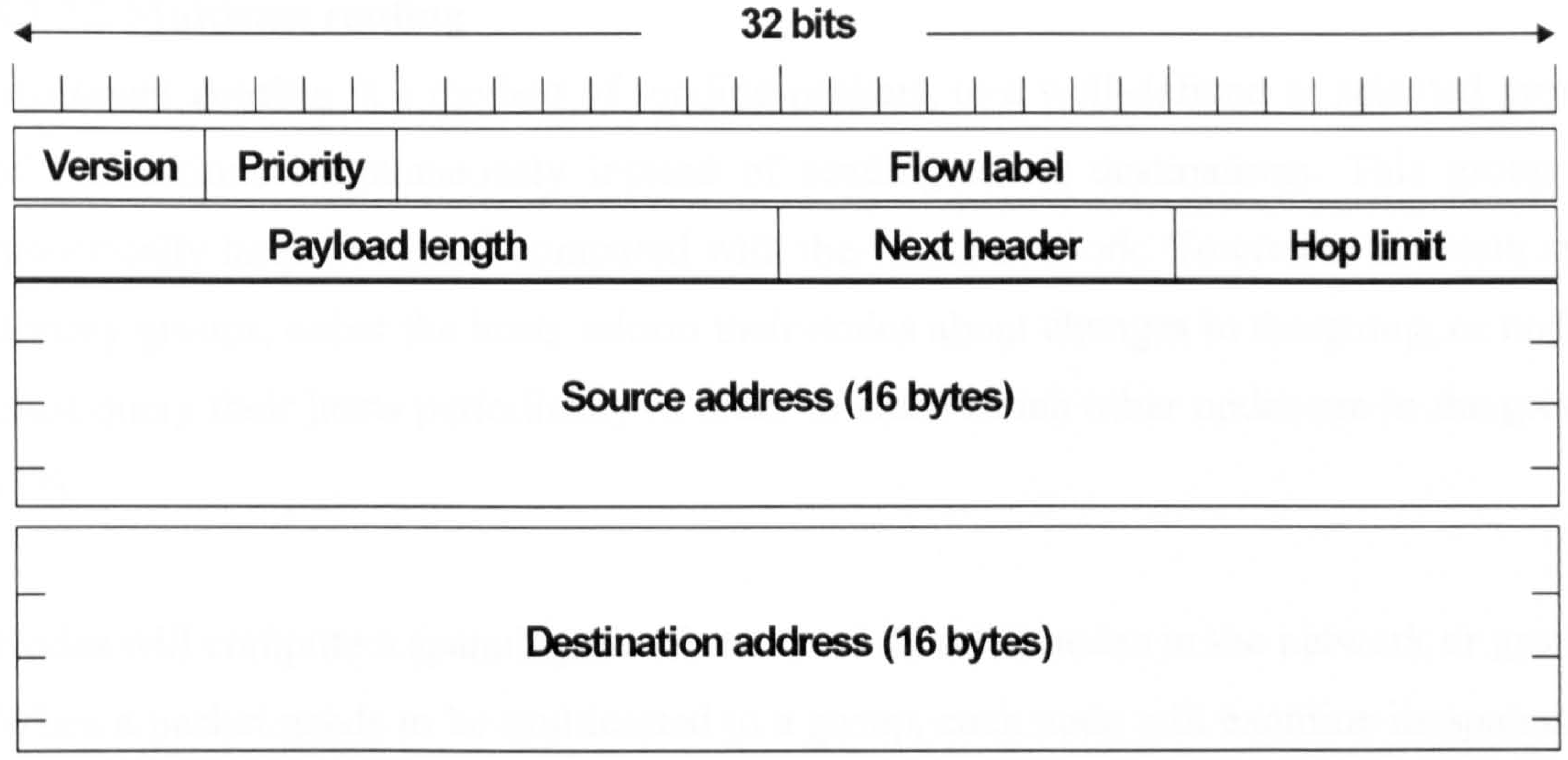
Broadcast routing is a method used for applications where packets need to be sent to unspecified destinations simultaneously, for example route update information, or online broadcast real time applications that can be shared by the public. Three methods can achieve packet broadcasting.

First is to send a packet to each destination simultaneously on predetermined routes. However, this method requires a complete list of routes leading to all destinations, and requires a lot of bandwidth.

The second is ***flooding***, in which case the nodes do not need a complete list of routes. The problem with this method is routers may generate too many duplicates and consume too much bandwidth. One of the most fundamental and simple methods for propagating routing information and packets is flooding, in which it forwards information without predetermined routes. Each node simply copies the received packets and forwards copies to all its neighbour nodes, except the node where the packet from. Such a way of propagation without control is very simple, but poses an infinite number of duplicated packets and stale routing information travels inside a network, causing confusion between the nodes, and triggers routing loops.

To prevent these problems, a *hop count* can be used as the packet's residual lifetime counter, to limit the distance that the packet can travel, and source nodes can attach a *sequence number* to an individual packet, so that other nodes can read this number and record the highest seen sequence number to later determine if a packet from the same source has been forwarded before or carries stale information.

Figure 15 (a) illustrates the Internet Protocol version 6 (IPv6) addressed packet header, containing a *hop limit* field, which is a hop count field with a maximum initial limit (e.g. 15 hops) and decreases at each relaying hop. Figure 15 (b) shows a sequence number entry recorded in node B's routing table, where only a packet from a source (A, C, D, E, F, G or H) with the highest sequence number will be used to update and overwrite routing information for the record of source, otherwise it is discarded as a stale duplication. For example, B has received 12 packets from A, 1 packet from C, 101 packets from F.



(a)

A	B	C	D	E	F	G	H
00012	▨	00001	00025	00060	00101	00053	00089

(b)

Figure 15 IPV6 addressed Packet

(a) Hop limit field (b) Sequence number table in router "B"

Flooding has the advantageous feature of being simple, is robust, has a short delay, and updates all the databases concurrently selecting every possible path in parallel, which could be beneficial in distributed database applications. However the exponentially increasing packet duplication exacerbates routing overheads. Many modified versions of flooding have been suggested in different routing algorithms developed over the years.

The third method is ***multi-destination routing***, which uses the packets themselves to carry a complete list of routes instead of storing them, as in a router. Each packet contains a list of destinations, the router checks all the destinations and determines the set of outgoing lines needed for transmission. This method is less demanding on the nodes, since it generates a large number of packets.

3.3.3.2 Multicast routing

Multicast routing is a method of sending packets, to a well-defined or selected group of destinations simultaneously instead of sending to all destinations. This group is numerically large but small compared with the whole network. To create, maintain and destroy groups, either the hosts inform their nodes about changes in the group, or nodes must query their hosts periodically in order to learn which other nodes are in the group [22].

Nodes will compute a spanning tree that covers all other nodes in the network or group. When a packet needs to be multicasted to a group, each node will examine its spanning tree of the network and mark different groups on the tree, then this packet will be sent only to those nodes that belong to the same specified group. Figure 16 (a) shows a map of the network with all routers represented by a node.

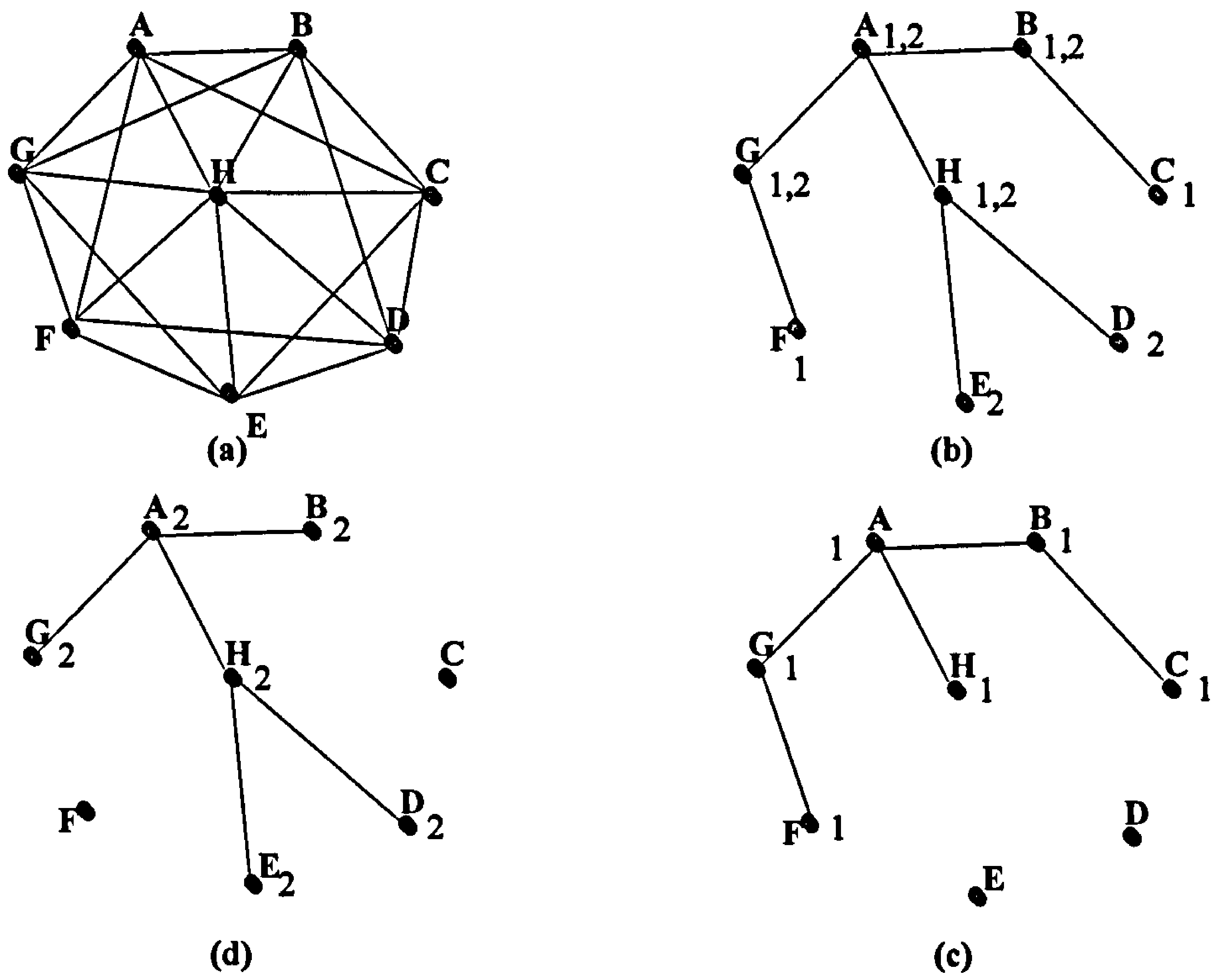


Figure 16 Multicasting network map and groups

(a) Network map (b) A's spanning tree (c) Multicast Group 1 (d) Multicast Group 2

Figure 16 (b) illustrates a spanning tree computed by node A, in which case all nodes are marked with a group ID to indicate which group they belong to. Figure 16 (c) shows all the nodes belonging to multicast group 1, including node A, B, G, H, E, and D. Figure 16 (d) is all nodes of group 2, including node A, B, G, H, F, and C. If a packet sent for group 1 is received by node A, it will forward this packet to all nodes on the spanning tree of group 1, but not to other nodes belonging to unrelated groups, as well as the multicast packet for group 2.

3.3.3.3 Unicast routing

Unicast routing is a method to propagate packets or information in a point-to-point fashion from a source to a unique destination [78, 79]. This method is commonly used in point-to-point communications on local area networks or Internets, where a connection is established for two terminals to communicate with each other. Unicast is one of the considerations of the IPv6 addressing standard. Moreover, protocols such as RIP and OSPF are designed to operate in unicast fashion. Table 3 shows a summary of the above three types of routing.

	Type of Transmission		
	Broadcast	**Multicast**	**Unicast**
Destination Addressed	Address to Unspecified destinations	Address to a group specified destinations	Address to a unique destination
Type of Communication	Public sharing communication	Group sharing communication	Paired user sharing communication
Method of routing	Sent to all members with predetermined routes; Flooding with control; Sent to multiple known destinations	Sent to multiple specified destinations	Point-to-point routing

Table 3 Different types of transmission used for Routing

3.3.4 Proactive (Table-Driven)

The routing algorithms introduced so far are designed and developed for centralized networks, whereas this section investigates routing algorithms that have evolved from these earlier schemes, for distributed routing in networks such as WANETs. Distributed routing algorithms can be classified, depending on how their routing table is constructed, into three basic categories: *proactive* known as table-driven routing algorithms, *reactive* or so-called on-demand source initiate algorithms, and *hierarchical* or hybrid routing algorithms, which combine proactive and reactive types of algorithms. This section carries out a brief overview of proactive routing algorithms.

The proactive type of routing algorithms predetermine and maintain a routing table, which records a track of routes connecting itself to all destinations in the network, in each node at all times, before the routing process actually starts. The advantage of proactive routing algorithms is: they have shorter initial route discovery delay, meaning that when a route is required, nodes can immediately select a predetermined route from the routing table without initiating route discovery each time.

However, proactive routing algorithms generate additional routing control traffic, due to each node wasting bandwidth to maintain routes even when it is not in use. The Destination Sequenced Distance Vector (DSDV) routing protocol [80], and the Optimised Link State Routing (OLSR) algorithm [75, 81] are the representative algorithms of this category.

3.3.4.1 Destination-Sequenced Distance Vector (DSDV) routing

DSDV is derived from the classic Distance Vector Routing (DVR) algorithm and is designed to suit dynamic network changes in WANET [1, 80]. DSDV preserves the simplicity of RIP [34], which is a well-known Internet routing protocol that fails to handle network changes in WANET, but at the same time avoids the routing loops caused by frequent route information exchanges. DSDV tags each routing table entry using a sequence number, so that nodes can distinguish stale routes from newly advertised routes, avoiding routing loops. The logical steps of routing operation used with DSDV can be summarized as follows:

- Each node maintains a routing table, in which are listed all known destinations and the distance (number of hops) required to reach them.
- Each route entry in the routing table is tagged with a sequence number.

- Each node broadcasts its own routing table to all its neighbours periodically or whenever there is a significant change of topology.
- If a route is unstable, broadcasting of a routing table could be delayed to reduce the number of rebroadcasts arriving with the same sequence number.
- A routing table can be updated by using its neighbours' routing tables, thus remaining consistent with the dynamically changing topology.

Figure 17 shows an example of a WANET topology map and the routing operation of node C using DSDV. Consider the node H moves from left to right on the map. The existing route in node C's routing table will change accordingly. Figure 17 (b) illustrates the routing table of C, which contains one entry for each destination before H moves. The contents of each entry are listed as follows:

- The Destination node address (Des.) indicates each available node in the network.
- The Next hop address (Next.) indicates the preferred neighbouring node to be used to forward a packet towards a particular destination.
- The metric (number of hops) indicates the measured distance to reach a destination.
- The Sequence number (Seq.) indicates sequence numbers originated by a destination for the new route leading to it.
- The Install time (Install.) indicates the time of deletion of a stale route.
- The Stable data (Stable) indicates a null structure, meaning no routes are to be superseded by be in competition with other possible routes.

(b)

Des.	Next	Metric	Seq.	Install	Stable
A	B	2	S280_A	T001_C	Ptr1_A
B	B	1	S218_B	T001_C	Ptr1_B
C	C	0	S654_C	T001_C	Ptr1_C
D	D	1	S107_D	T001_C	Ptr1_D
E	D	2	S923_E	T002_C	Ptr1_E
F	D	3	S607_F	T001_C	Ptr1_F
G	D	4	S812_G	T001_C	Ptr1_G
H	B	2	S500_H	T002_C	Ptr1_H

(c)

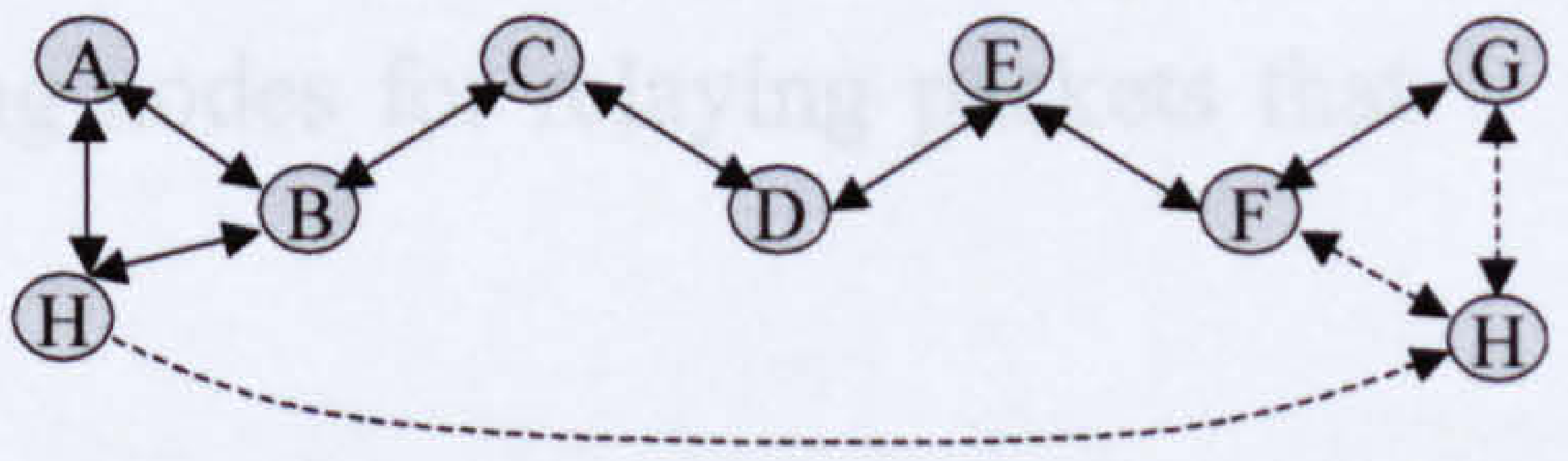
Des.	Next	Metric	Seq.	Install	Stable
A	B	2	S280_A	T001_C	Ptr1_A
B	B	1	S218_B	T001_C	Ptr1_B
C	C	0	S654_C	T001_C	Ptr1_C
D	D	1	S107_D	T001_C	Ptr1_D
E	D	2	S923_E	T002_C	Ptr1_E
F	D	3	S607_F	T001_C	Ptr1_F
G	D	4	S812_G	T001_C	Ptr1_G
H	D	4	S523_H	T032_C	Ptr1_H

(d)

Des.	Metric	Seq.
A	2	S280_A
B	1	S218_B
C	0	S654_C
D	1	S107_D
E	2	S923_E
F	3	S607_F
G	4	S812_G
H	2	S500_H

(e)

Des.	Metric	Seq.
A	2	S280_A
B	1	S218_B
C	0	S654_C
D	1	S107_D
E	2	S923_E
F	3	S607_F
G	4	S812_G
H	4	S523_H



(a)

Figure 17 DSDV Routing

[1, 80] (a) WANET topology map (b) C's routing table (c) Updated routing table of C (d) Broadcast C's routing table (e) Broadcast of updated C's routing table

Figure 17 (c) illustrates the updated routing table of node "C" after node "H" has moved from its left to the right of the network. The highlighted entry is the updated entry for node H. Figure 17 (d) is the routing table of node "C" broadcast to all its neighbours before node "H" has moved. Figure 17 (e) is the updated version of the routing table broadcast to all its neighbours after node "H" has moved and the updated entry for node H has been highlighted. The broadcast table includes the Destination node address (Des.), the "metric" (number of hops) and the Sequence number (Seq.). All neighbouring nodes of C receiving the updated routing table will update their own routing tables. The updating process occurs either periodically, or when a link is broken off or comes back on. To maintain consistency of the dynamic network topology, the rapid update of routes is necessary.

The advantages of DSDV are the simplicity, short initial route discovery delay, and periodical updates of route information according to topological changes. The drawback of DSDV is when a new node joins in, or an existing link breaks off, it will broadcasting of connectivity information, which wastes network resources as it maintains all paths leading to all destinations at all times, even for routes not presently needed. The packet overhead will increase exponentially as the number of nodes in the network grows larger. The proposed Optimised Link State Routing algorithm (OLSR) can thus minimise part of the negative impact of DSDV.

3.3.4.2 Optimised Link State Routing

OLSR is the optimization of the classic LSR introduced earlier. It is different from DSDV, in the sense that it only maintains those routes needed. OLSR is also different from LSR, since it uses a set of selected neighbouring nodes for relaying packets that are send from any node [1, 81].

Any source node (e.g. S_1 shown in Figure 18) selects a set of relaying nodes among its one-hop neighbours, known as the Multipoint Relay (MPR) set (shown as the grey dots in Figure 18). The MPR nodes must cover all two-hop nodes and only a member of this MPR set may retransmit packets from the corresponding working source node. The working node can also be selected as a member of another node's MPR set, and that neighbouring node is called a Multipoint Relay Selector (MPRS as S_2 shown in Figure 18).

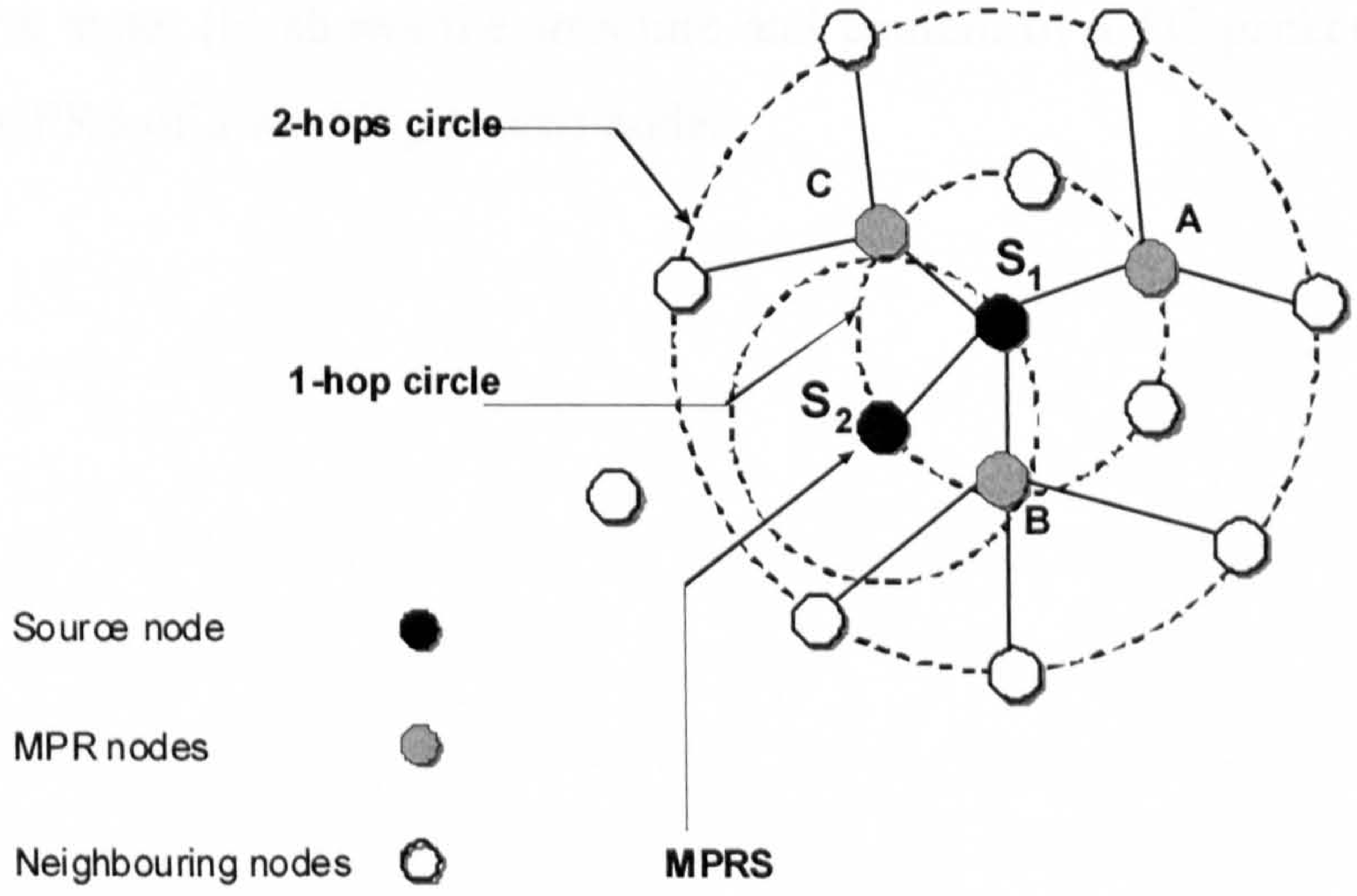


Figure 18 Multipoint Relay mechanism of OLSR [82]

The source node records the MPR in a HELLO packet, which is broadcast periodically. If a node receives a HELLO packet from a neighbour, and if this node's ID is in the MPR set that comes from that neighbour, it will list the source node of the HELLO packet as an MPRS.

Each node learns those MPRS surrounding it and generates a Topology Control (TC) packet, then broadcasts this to all other nodes in the network, only via their MPRs, so that other nodes can update their topology map [82]. In this way the route information in TC packets can be broadcast to the entire network, whilst significantly reducing duplicated retransmissions.

Each node has a Multipoint Relay Selector Sequence Number (MSSN), which is increased whenever the MPR list of this node is updated. Each node relies on MPRs to obtain and propagate routing information, then calculates or updates the shortest paths in its routing table. Each node periodically broadcasts the route information detailing which neighbours have selected it as an MPR. The route is a sequence of hops through the MPRs from the source to the destination [1, 82].

Figure 19 illustrates two different routing information packets used by OLSR: (a) shows the HELLO message, which records an MPR node set with their addresses and their link state; (b) shows the structure and content of a TC packet, which records the list of MPRS of a working source node.

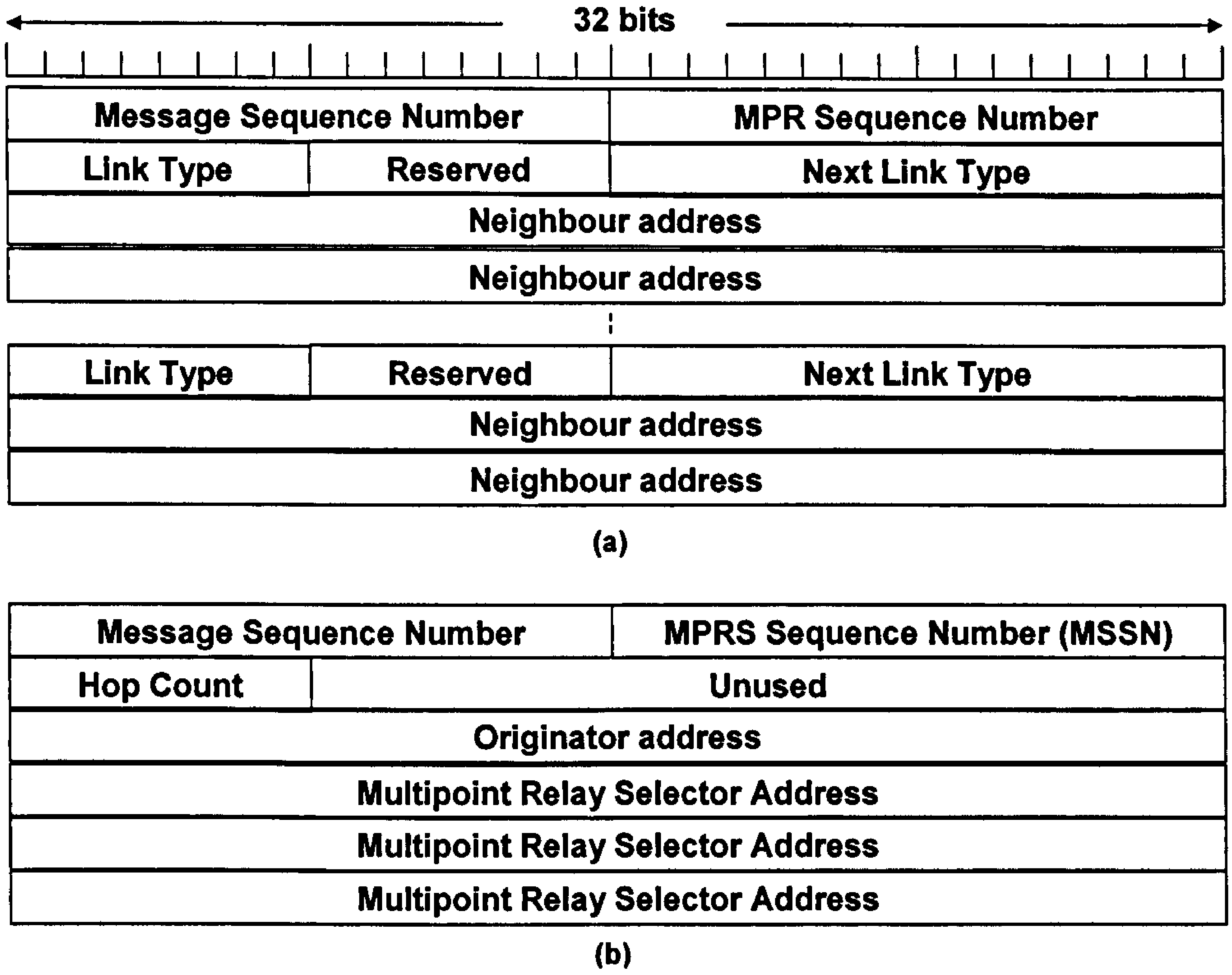


Figure 19 OLSR route information packets [1, 81, 82]

(a) Hello message (b) Topology control message

OLSR has made additional improvements in comparison to the classic LSR. Firstly, it reduces the size of the control packets, so each packet only records information of MPR set nodes. Secondly, OLSR uses MPRs to minimise the control packets (similar to the link state packets in LSR) flooded in the network and reduces the routing control traffic caused by packet retransmission. OLSR simply reduces the time interval for periodic control packet transmission, so that the route information stays consistent with network topology changes [1, 81, 82].

The major drawback of OLSR is the proactive nature of routing, which means network resources are wasted due to predetermined routes leading to all destinations being maintained at all times, even for routes not currently in use. To solve this problem, reactive or source initiate on-demand type of routing algorithms are proposed.

3.3.5 Reactive (Source-initiated/On-Demand)

Reactive routing algorithms activate a route discovery procedure by source nodes when a route is needed. This results in more efficient use of network resources owing to the fact that no maintenance is required for routes that are not in use. The route discovery process is complete once a route is found, or all possible routes have been examined, and the established route is maintained until the destination becomes inaccessible via every path or the route is no longer required [1, 9, 13].

The advantage of reactive routing algorithms is that the bandwidth consumption for maintaining the a routing table in each node is far less than proactive algorithms and it is loop free. On the other hand, reactive algorithms are slow to converge in initial route discovery, which causes long initial delay before the real data communication starts. Dynamic Source Routing (DSR) [83], ad hoc on-demand distance vector (AODV) [84], temporally ordered routing algorithm (TORA) [85], and the Associativity Based Routing (ABR) [86] belong to this category.

3.3.5.1 Dynamic Source Routing (DSR)

DSR is designed for multi-hop WANET routing, and is composed of two mechanisms, route discovery and route maintenance. DSR uses source routing, which obviates the use of intermediate nodes to forward routing packets and obtains up-to-date routing information. DSR accesses and propagates routing information that is piggybacked in a data packet header instead of advertising routing information periodically, or sensing link status, or using neighbour detection packets. DSR allows the routing overhead in the data packet to be scaled to only what is needed for reacting to changes on the currently used routes [83].

DSR can discover single or multiple routes in a one-route discovery operation. Each packet carries in its header the complete list of intermediate nodes that it must pass through. Routing information is cached and updated in intermediate nodes when forwarding or overhearing packets. DSR avoids initiating a new route discovery when a source route fails and instead uses alternative routes previously discovered for the same destination. DSR records routes that have been discovered previously in a route cache that is equivalent to a routing table in each node.

DSR supports unidirectional routes and asymmetric links, in which case a link between two nodes may not work equally well in both directions due to different antenna functions, propagation patterns, or transmission power levels. In the case where reverse links are not available on some part of a route, DSR allows nodes to send packets to any other nodes using multiple links. This is a major advantage for a reactive routing algorithms that were later proven to be very useful in developing our own original routing strategy called *Minimum Impact Routing* (MIR), which will be discussed in more detail in chapter 5-7.

In a single route discovery, the initiator can learn and cache multiple routes to any particular destination. In this way, if one source route is changed, the initiator can still use other routes leading to the destination. This increases the capacity for rapid reaction to route changes and avoids the increasing packet overheads caused by initiating a new route discovery procedure whenever a route is broken.

Figure 20 illustrates the DSR mechanisms, where (a) shows a Route Request (RREQ) packet with an ID number (2) originating from initiator A (Sour. in the packet), and records intermediate nodes (Int. in the packet) along one of the paths (A-B-C-E-H-I) to the target node I (Des. in the packet).

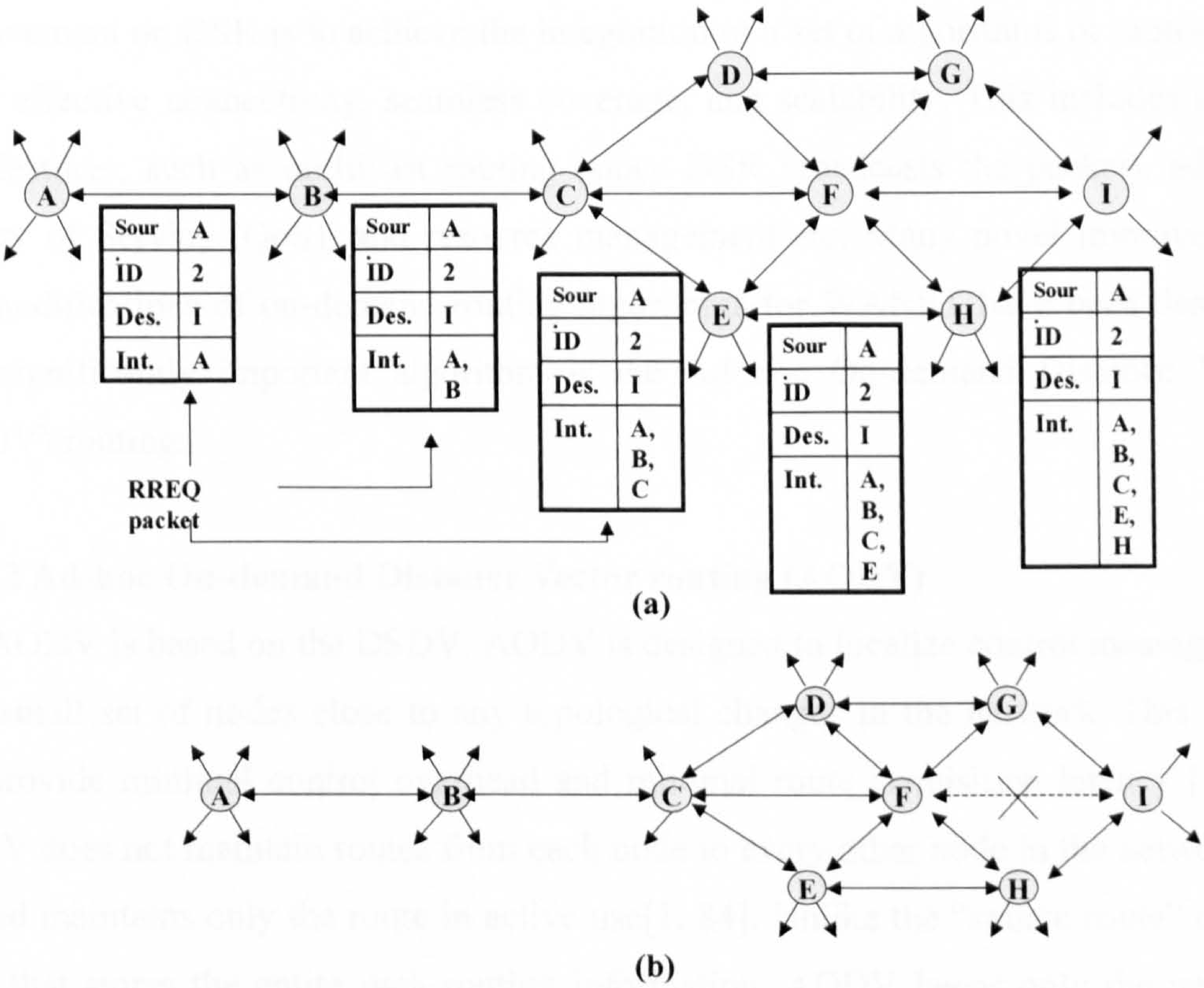


Figure 20 DSR routing mechanisms

(a) Route discovery (b) Route maintenance

The intermediate node address field accumulates the addresses of each node along this path, increasing as nodes along this path forward the RREQ packet. Figure 20 (b) illustrates route maintenance when one link along the source route is broken. A packet is sent from source node A towards destination node I along a known multi-hop path A-B-C-F-I. Along this path, node B will be responsible for sending acknowledgement back to A, telling A it has received the packet, and respectively C acknowledges to B, F to C, I to F and so on. Assuming the link between node I and F is not available, F may send the packet to I via G or H, and returns an Route Error (RER) message to node A. When node A receives the RER, it scans the local route cache, sending out the packet using an alternative route to the same target I.

DSR is a simple and effective algorithm, by which a source node can learn routes to each intermediate node on the "source route" by using a single request and reply cycle in the route discovery process. All intermediate nodes can learn the route to other nodes in the network, as each node stores all source routes learned in the "route cache". The improvement on DSR is to achieve the integration of a set of algorithms or protocols to allow effective connectivity, seamless coverage, and scalability. This includes adding new features, such as multicast routing (since DSR broadcasts the packet), adaptive Quality of Service (QoS), and resource management etc. Many novel improvements and modifications of on-demand routing algorithms for WANET have been designed. One significantly important algorithm is the Ad hoc On-demand Distance Vector (AODV) routing.

3.3.5.2 Ad-hoc On-demand Distance Vector routing (AODV)

The AODV is based on the DSDV. AODV is designed to localize control messages to a very small set of nodes close to any topological changes in the network. This design can provide minimal control overhead and minimal route acquisition latency [1, 84]. AODV does not maintain routes from each node to every other node in the network but instead maintains only the route in active use[1, 84]. Unlike the "source route" used in DSR that stores the entire path routing information, AODV keeps only the next hop routing information. It also reduces the number of broadcasts triggered by link breaks; if the link does not affect ongoing communication or actively maintained route, no broadcast of routing information occurs.

AODV supports three types of communication: unicasting, multicasting, and broadcasting. For unicast and multicast routing, each method has two stages of operation: route discovery and route maintenance[1, 84]. Routes are discovered on an on-demand basis and maintained only as long as they are needed.

Route establishment is associated with aging, i.e. each route is established for a source node that exists only within its lifetime[1, 84]. This means nodes do not store unnecessary route information in order to maintain a route that is not being used. The aging of "route packets" prevents a network from wasting system resources through exchanging stale route information. It also prevents confusion in routing operations caused by receipt of stale route information. The aging of a route entry is also applied in AODV. Each node maintains a routing table, which consists of one route entry for each destination node. The aging of a route entry prevents a node from using stale routes to transmit packets to other nodes after they have moved from their previous position. Whenever a route is not used, its route entry in the routing table will expire and be discarded [1, 84].

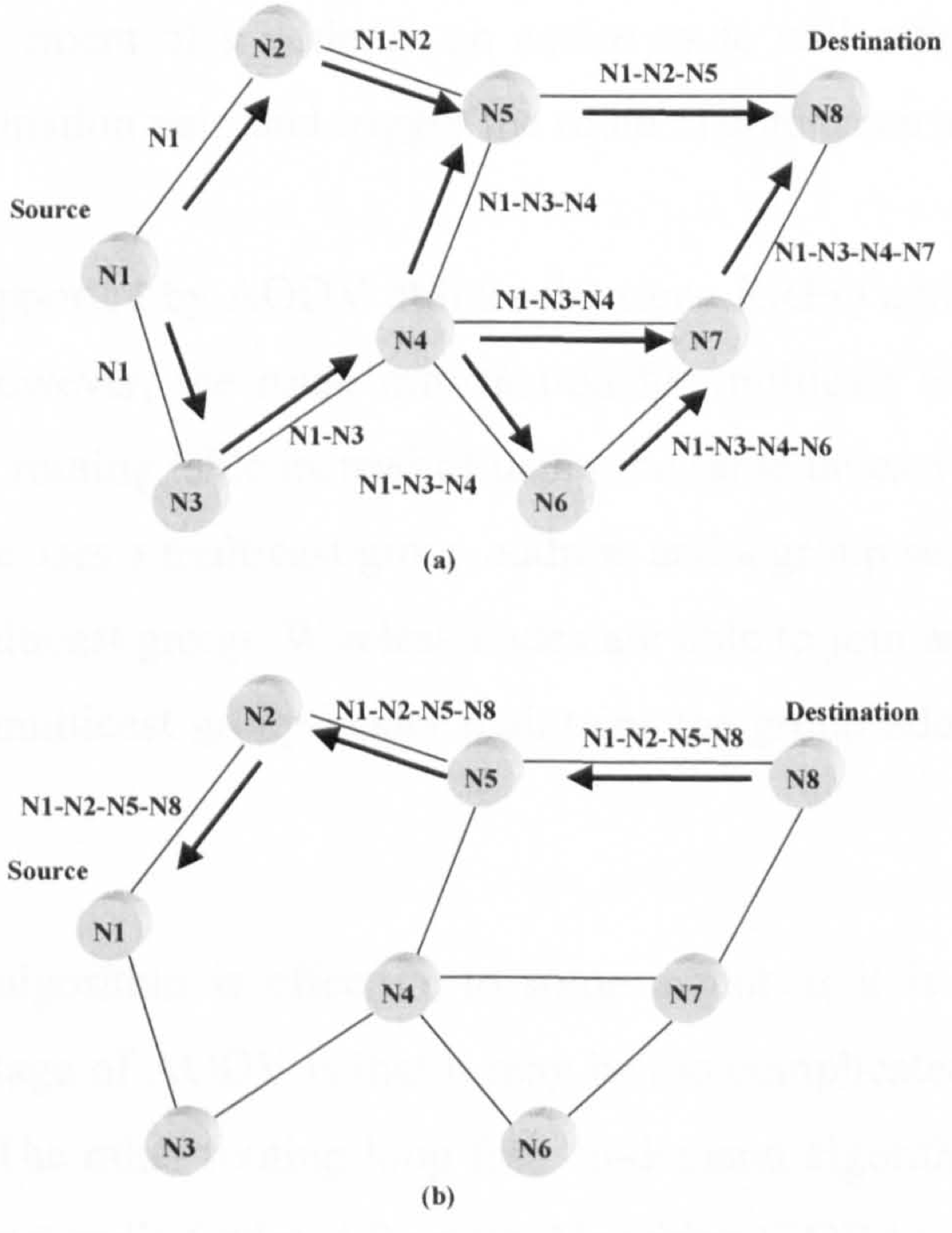


Figure 21 AODV route discovery process

(a) route request (b) route reply [1, 84]

AODV supports both point-to-point and point-to-multipoint communications. For point-to-point communication, AODV uses unicast routing, and for point-to-multipoint communication, it uses multicast routing. The route information obtained when searching a multicast route can also be used for updating unicast routing knowledge. Figure 21 illustrates the route request and reply process in AODV, where (a) shows the route request process, and (b) is the route reply process. The following review of AODV consists of two parts: *unicast* and *multicast* routing.

Unicast routing in AODV follows a request-reply cycle. AODV utilises source and destination sequence number to ensure that the routing process is loop-free, and the routing table only keeps the most recent route information. Once a source node starts *a route discovery* process, it creates a Route Request (RREQ) packet, and broadcasts the RREQ packet, at the same time setting a timer to wait for the reply. If no Request Reply (RREP) is received before the timer expires, either the RREQ is not being correctly transmitted or the destination has not been reached before the RREQ's lifetime runs out, and the source node will broadcast a RREQ retry procedure, with a longer lifetime. The *route maintenance* procedure is used to maintain an existing route as long as it is needed. Only the movement of a node on an active route will affect communication between a source-destination pair, and trigger the route maintenance procedure.

Multicast routing supported by AODV utilizes the same RREQ and RREP cycle used in unicast routing. However, the route information for multicast routing is stored in each node's multicast routing table instead of using the same unicast routing table. The multicast routing table uses a multicast group address and a group sequence number for all nodes within a multicast group. Wireless nodes are able to join and leave the group at any time, and the multicast group leader maintains the group address and sequence number.

The AODV routing algorithm is effective to some extent as it is routing loop free. However, a disadvantage of AODV is that it may be too complicated to implement for certain applications. The other routing loop free on-demand algorithm uses a different approach, called a Temporally Ordered Routing Algorithm (TORA).

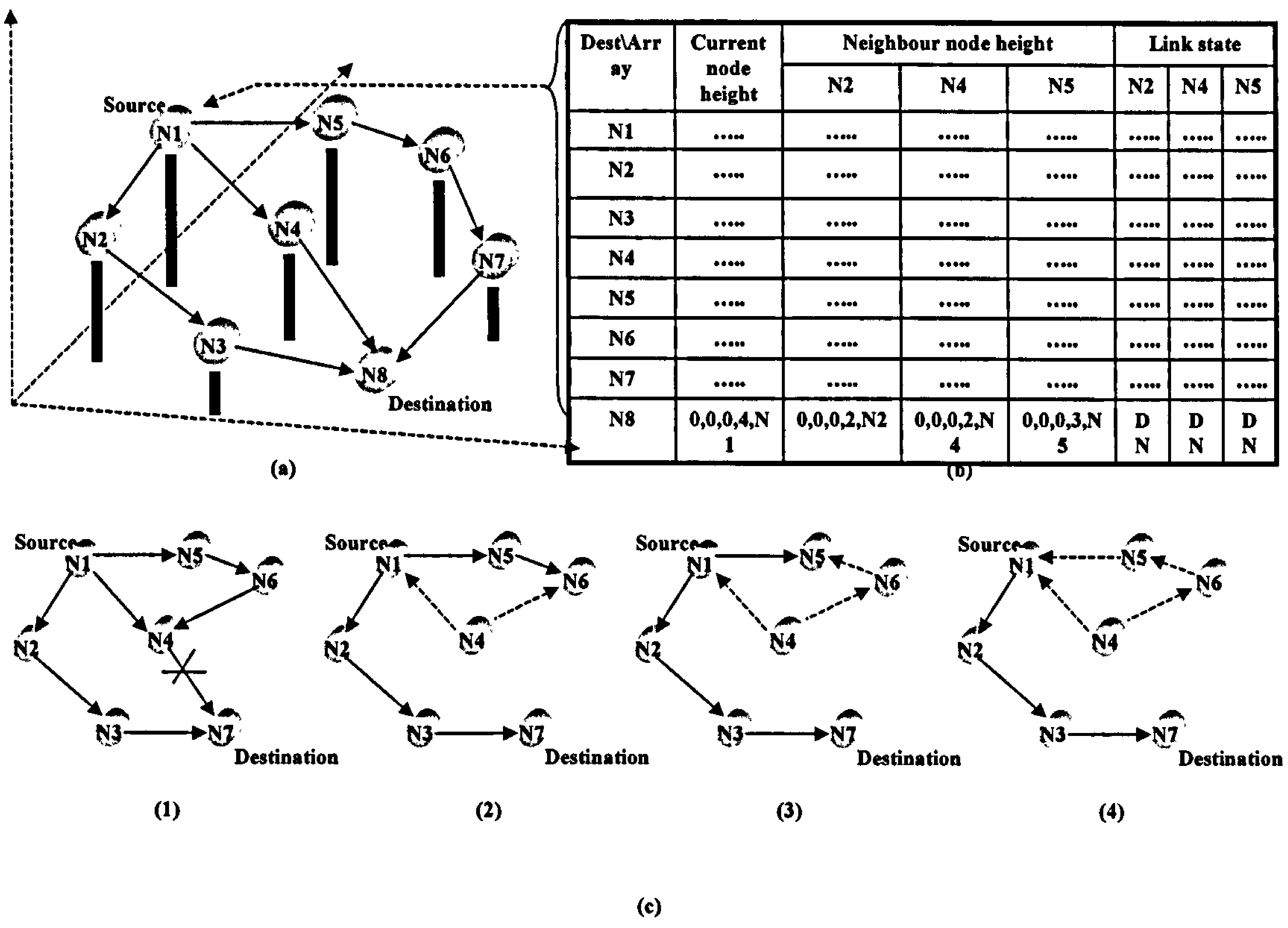
3.3.5.3 Temporally Ordered Routing Algorithm (TORA)

TORA is a highly adaptive, loop-free, distributed routing algorithm. The concept of this algorithm is based on link reversal routing (LRR) [87, 88]. Three representative algorithms are developed base on LRR, the Gafni-Bertsekas [89] algorithm, the Lightweight Mobile Routing (LMR) [90] algorithm, and TORA [1, 87, 88], where all these algorithms are based on the concept of "link reversal", but differ in their routing organization. Instead of maintaining a distributed network state to compute a shortest-path, LRR maintains only sufficient state information to constitute a Directed Acyclic Graph (DAG). For each source – destination pair, LRR gives each known node a reference level measured by a "height metric", LRR uses this reference level construct a DAG temporarily.

TORA combines features of both GB and LMR, and the following description will concentrate TORA. It utilizes the "request-reply" mechanism and the "partial link reversal" mechanism, in order to exchange control information and maintain routes. The routing information is carried in three types of control packets: Query (QRY), Update (UPD) and Clear (CLR). The routing used by TORA can be briefly explained in three basic functions: route creation, maintenance, and erasure.

Route creation of TORA measures the height array of each of its neighbours, and determines link direction to construct a DAG, which contains multiple routes leading to a destination node in a previously undirected network.

Figure 22 (a) shows a DAG with three routes starting from the source node (N1) to the destination node (N8) via different intermediate nodes, and in (b) a height array list indexed by destination node IDs, the reverse order depends on the height of each node. Figure 22 (c) illustrates the first four steps of directed link reversal. The upstream node of the broken link reverses its incoming links by broadcasting an UPD packet to its neighbours, and this operation continues to reverse all related links in the DAG.



Dest\Array	Current node height	Neighbour node height			Link state		
		N2	N4	N5	N2	N4	N5
N1	…..	…..	…..	…..	…..	…..	…..
N2	…..	…..	…..	…..	…..	…..	…..
N3	…..	…..	…..	…..	…..	…..	…..
N4	…..	…..	…..	…..	…..	…..	…..
N5	…..	…..	…..	…..	…..	…..	…..
N6	…..	…..	…..	…..	…..	…..	…..
N7	…..	…..	…..	…..	…..	…..	…..
N8	0,0,0,4,N1	0,0,0,2,N2	0,0,0,2,N4	0,0,0,3,N5	DN	DN	DN

Figure 22 TORA routing mechanism

(a) DAG for node N1to N8 (b) Height Array of N1 (c) First four steps of "link revesal" [1, 87, 88]

TORA route maintenance process starts after any node in the DAG has lost its last outgoing link. It removes its height and link status metrics from them, and then reselects their heights and reorients the DAG, all directed paths would then lead to the destination again after a DAG reconstruction.

Figure 23 illustrates 5 cases, which will start the maintenance procedure in a logical top-down diagram. All directed paths then will lead to the destination again after a DAG reconstruction. *TORA route erasure* takes place if a small group of nodes find themselves apart from other nodes in the network, this situation is called network partition, and a route erasure procedure will start.

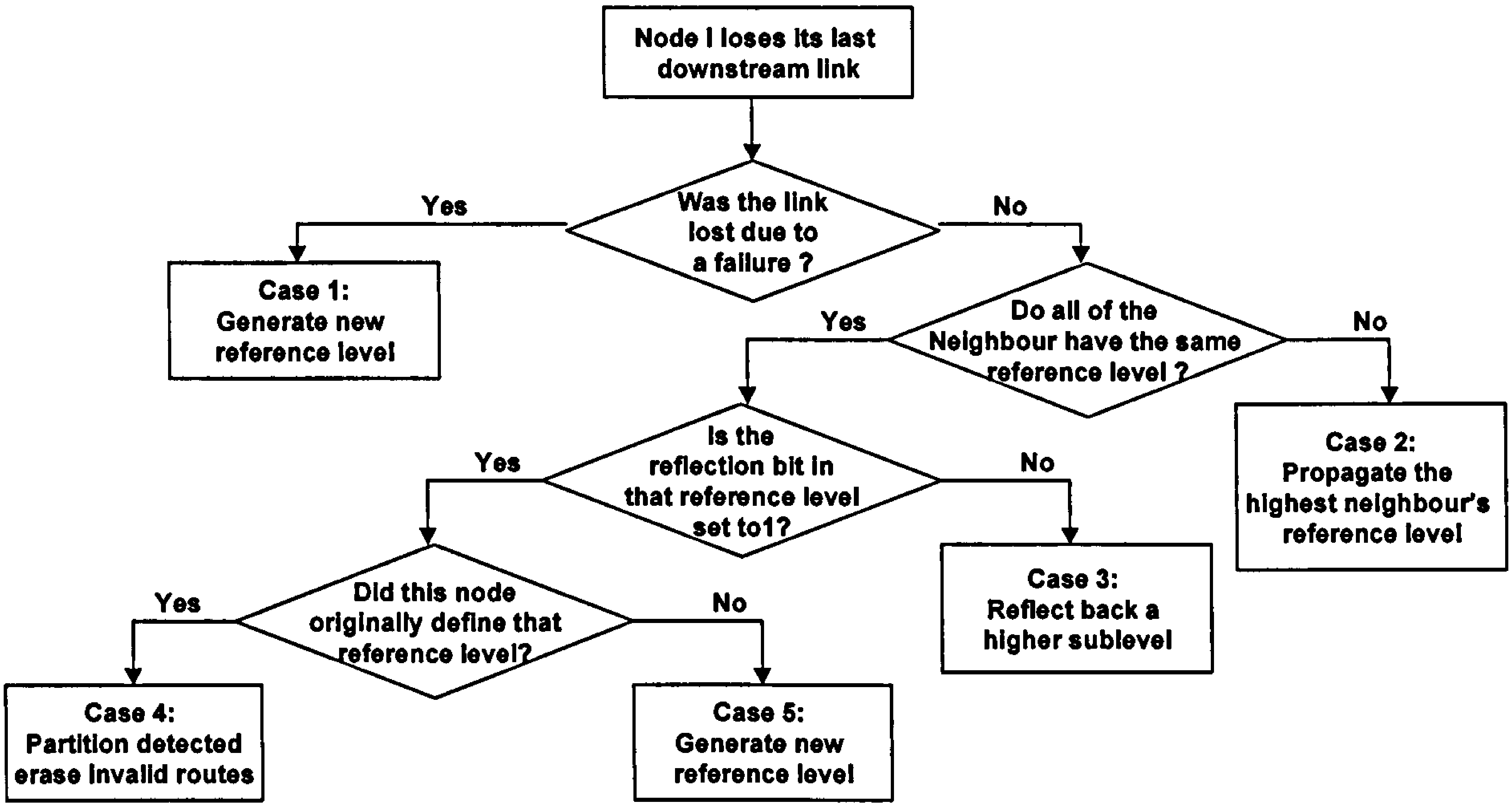


Figure 23 Route maintenance procedure [1, 87, 88]

The advantage of TORA is minimization of the amount of routing overhead exchanged between nodes by localizing exchanges of control messages to a very small set of nodes near topological changes. TORA maintains only sufficient routing information to constitute a DAG, which is rooted at the destination node. TORA can be described as "topological change-oriented", which means maintaining routes only in response to topological changes. The drawback of TORA is the use of a physical or logical clock, in order to achieve the synchronization of building a DAG temporally, and there will be a DAG for each source-destination pair.

3.3.5.4 Associativity Based Routing (ABR)

Another on-demand algorithm developed by C-K Toh and Vasos is called Associativity Based Routing (ABR) [86], which is a source-initiate/on-demand that considers associativity of neighbouring nodes, and emphasises associativity of mobile hosts (referred to as nodes in the following) in a dynamic network environment such as

WANET. Associativity is measured by a node's connectivity relationship with its neighbours, which will change when a node moves across its neighbour's wireless cell/coverage area.

This transition period is measured by an associativity ticks ($A_{threshold}$) [86]. For example, a wireless cell size is about 10 meters (m) in diameter, and a node moves at a speed of 2 m/sec across a wireless cell. One or more neighbouring nodes will record an $A_{threshold}$ of no more than 5 seconds, any $A_{threshold}$ greater than 5 indicating a period of association stability. A node with a high associativity (i.e. greater than $A_{threshold}$) means, neighbouring nodes surrounded the node, and its connectivity with them is stable. If a node has a high associativity, and is in a stable state, it has low mobility. On the other hand, a lower associativity means high mobility, and an unstable state.

During a routing process, all nodes can be categorised into three types: Source node (SRC), Destination node (DEST) and Intermediate node (INT). Each node periodically broadcasts beacons (like a hello message) to identify itself and learns its neighbouring nodes. Each node constantly updates its associativity ticks in accordance with other neighbouring nodes. ABR employs a Broadcast-Query (BQ) and Await-Reply (REP) cycle to accomplish route establishment and maintenance. When a route is no longer needed for a source node, it will delete the route by sending a Clear (CLR) message. The mechanism of ABR can be described in three phases: route discovery, route reconstruction and route deletion.

Route discovery starts when a SRC needs a route to communicate with a DEST. SRC broadcasts a BQ packet with few packet header fields. These fields include a unique Sequence Number (SEQ), a source address/identifier (ID), destination address/ID and a hop count. Any node receiving this BQ packet, will check the sequence number of it. If the BQ from the same source with a greater sequence number has been recorded previously, the new arrived BQ with a less fresh sequence number will be discarded. If the new arrived BQ with a greater sequence number, the node will update the route entry in its routing table use this new BQ.

Route reconstruction (RRC) procedure will be activated when the association stability relationship is violated. ABR attempts to manage link break up by relocating an alternative valid route instead of restarting a route query broadcast. Assuming that the direction towards the destination node is downstream and that the direction towards the source node is upstream, the RRC may consist of four different types of operations, depending on which node is along the route move. These four types of operations are as follows: new route discovery, partial route discovery, invalid route erasure, and valid route update.

Route deletion will begin when an SRC no longer desires a route. A route Delete (RD) packet broadcast will be initiated by the SRC. In this event all INTs will update their routing table entries for this SRC and delete the route announced in the clear up. A route deletion may be accomplished either by broadcasting RD packets, or just time out of the route entry when no traffic is related to the route over a period of time.

The drawback of the ABR is that when the system prepares to shutdown, it actually uses more power than usual, which does not meet the design goal of reducing the power consumption. Both the proactive routing and the reactive routing algorithms have some drawbacks and some advantages. Neither pure proactive nor reactive routing can perfectly manage a high mobility network with frequent changes in topology, nevertheless hierarchical routing combines the benefits of the two, to accommodate drawbacks of each type to some extent.

3.3.6 Hierarchical (Hybrid)

Hierarchical/hybrid routing algorithms reduce routing control traffic, packet overheads, improve routing efficiency, network capacity, and throughput. This type or routing uses a two level hierarchy to organize routing, where the first level groups a small number of routers into a cluster/zone, and the second level connects a number of clusters into a subnet.

Each cluster has one cluster head, which is a node responsible of resource allocation etc. Each node will constantly store routing information only for other nodes in the same cluster. In this case the amount of routing information stored in each router is much smaller. Each cluster contains one node acting as *cluster head*, which is responsible for scheduling transmissions and resource allocation. Each cluster has one or more *peripheral nodes*, within the radius of a cluster head, and zero or more *ordinary nodes*, which are neither a cluster head nor a peripheral node.

This section will introduce two hybrid routing algorithms: the Clusterhead Gateway Switch Routing (CGSR) protocol [91], and the Zone Routing Protocol (ZRP) [92]. CGSR and ZRP are distinguished by different clustering algorithms, which consist of *cluster formation* and *clusterhead selection* methods. CGSR groups nodes into a cluster, and then elects a node to act as the clusterhead, whilst ZRP forms a cluster for each node, and this inspires the later development of MIR in this research. The following sections will describe the clustering algorithms and scheduling algorithms of CGSR and ZRP.

3.3.6.1 Clusterhead Gateway Switch Routing (CGSR)

The main feature of CGSR is that the network is organised on a cluster-based structure, and transmission between nodes is scheduled in a contention-free manner. Transmission to adjacent clusters can be isolated through spread-spectrum multiple access schemes with each cluster using different spread codes [91]. CGSR routing algorithm can be summarized into three phases: clustering, gateway switching, and routing.

Clustering in CGSR could be identifier-based [93, 94] or connectivity-based [95], in which both serves two objectives: the *selection of clusterheads* and *formation of cluster*. CGSR uses different spreading codes (i.e. TDMA or CDMA) to achieve *spatial reuse* across clusters. To communicate with members in the same cluster, the clusterhead uses a "controlled transmit permission token protocol" (i.e. polling) to allocate the channel for transmission. This allows clusterheads to prioritise the transmission of packets in its queue.

Gateway switching in CGSR is carried out by a peripheral node, which belongs to more than one cluster or is linked to a gateway node in another cluster. To communicate with a node in an adjacent cluster, the gateway node needs to switch from one set of codes to another. Conflicts may occur in terms of loss of the permission token for transmission when gateway nodes tune to another code.

Figure 24 illustrates a subnet using CGSR, and mobile nodes organized in a link-clustered control structure. In this subnet, three clusters are formed. Cluster 2 and 3 overlap each over, and one gateway node is directly linked with the two cluster heads.

Cluster 1 is disjoined from the other two clusters, with two gateway nodes indirectly connecting the clusterhead in cluster 1 with the two cluster heads in clusters 2 and 3.

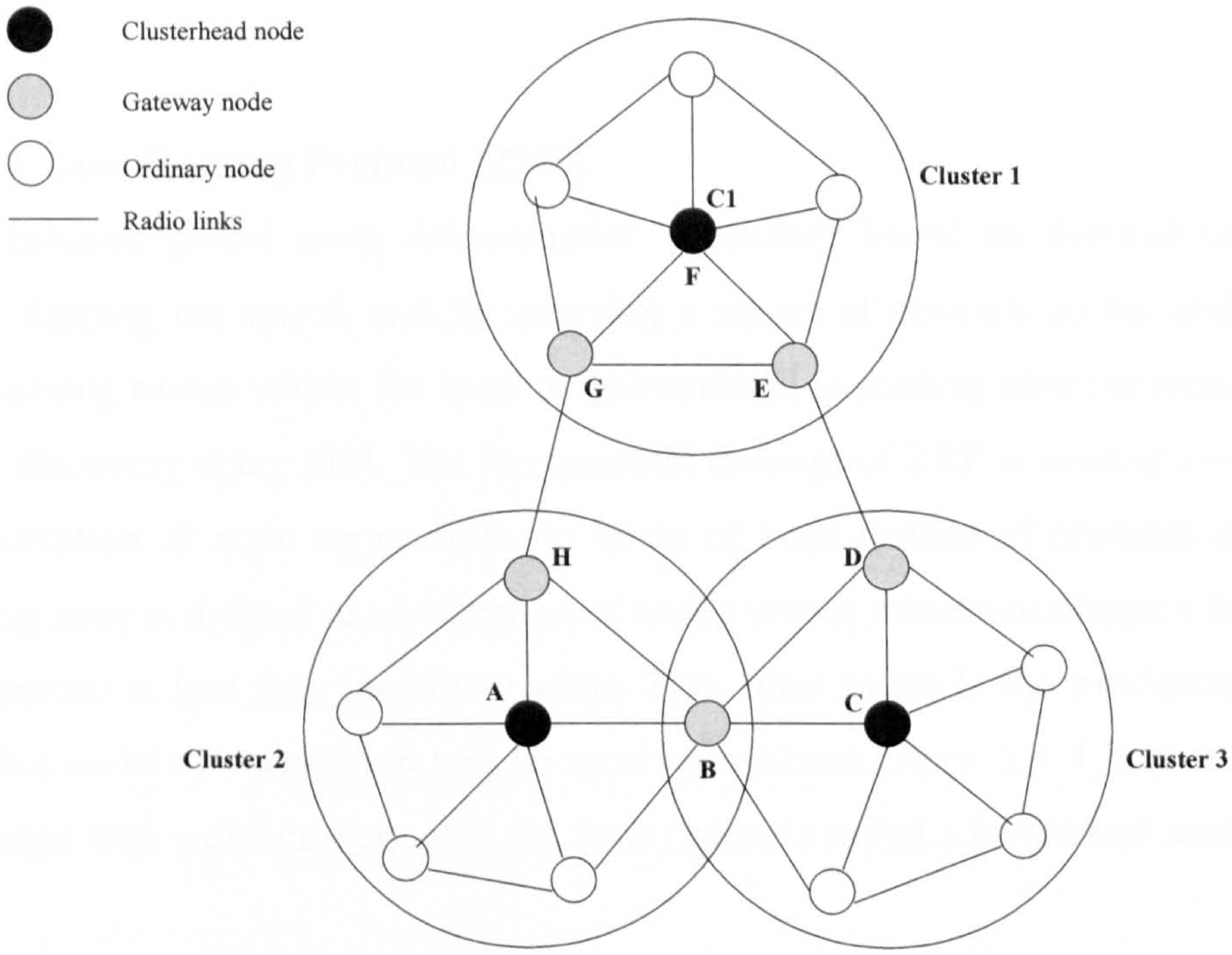


Figure 24 CGSR link-clustered control structure [91]

Routing in CGSR combines scheduling and hybrid routing, as shown in Figure 24, transmitting a packet from node A (the clusterhead of cluster 2) to node C (the clusterhead of cluster 3) via node B (the gateway).

CGSR uses the hybrid clusterhead-to-gateway routing approach to route traffic, which means an ordinary node will forward all its packets to the clusterhead node, and connect to another clusterhead node via gateway nodes.

The drawback of CGSR is that the clusterhead acts as a point of concentrated traffic, which may become congested, and so, a point of failure. In addition, because CGSR uses DSDV as the underlying routing scheme, it has the same overhead as DSDV. The proactive routing algorithms (i.e. DSDV) are not applicable, since they continuously use significant network capacity to keep the routing information current (quick on route discovery, but wastes network resources on maintenance). However, the reactive routing algorithms (e.g. DSR, AODV etc.) also need a global route search procedure, and this causes significant control traffic during searching (slow on route discovery, and costly network resources when searching for a route). The ZRP is another hybrid routing algorithm proposed to improve the robustness of routing and to reduce the overhead.

3.3.6.2 Zone Routing Protocol (ZRP)

ZRP initiates global route determination procedure based on demand of the source node, limiting the search cost by querying a subset of network nodes, and proactively maintaining routes within the local neighbourhood (a routing zone) to reduce the initial route discovery delay [92]. The fundamental concept of ZRP is *routing zones*, which is a description of node connectivity in terms of hops instead of physical distance. The routing zone is defined as a collection of nodes whose minimum distance from *the node in question* is less than the *zone radius*. This zone radius is the maximum distance at which a node's transmission will be received without errors (i.e. 1 hop or 2 hops), and any node with a distance equal to the zone radius is called a *peripheral node*.

The Neighbour Discovery Protocol (NDP) in ZRP is responsible for discovering neighbour nodes and constructing zones for each node. The Intra-zone Routing Protocol (IERP) is responsible for proactively discovering and maintaining routes for destinations inside a routing zone. The IERP is responsible for reactively discovering routes leading to destinations outside local routing zones.

During the discovery and maintenance of routes, ZRP produces control traffic, which includes intra-zone route update packets and inter-zone route request/reply/failure packets. ZRP generates control overhead when exchanging beacons. Therefore the performance of ZRP is mostly affected by control traffic and the control overhead. However, ZRP produces less control traffic than purely proactive routing algorithms (i.e. DVR or LSR) by using IARP to limit proactive routing within a routing zone. The amount of IARP control traffic required to maintain a routing zone increases with the routing zone size.

The drawback of ZRP is that when the network is relatively small, and wireless nodes are less active, the instantaneous network load will be dominated by the control traffic from a single route discovery, in which case, congestion may frequently occur. There is also an argument that the node acting as a clusterhead will consume more power to process additional tasks, such as administration or resource allocation and so on.

3.4 Asymmetrical Routing

Most of the current work on mobile ad hoc networks assumes symmetric connective in both directions, i.e. if node A can communicate directly with node B, then node B must be able to communicate directly with node A. In practice this will not be the case, due to different node capabilities (e.g. limited or variable transmit power, or limited battery life), or because of excessive localised interference requiring higher received powers in order to achieve an adequate SINR (reducing the maximum link lengths that can be supported). In AODV, each node receiving a Route Request (RREQ) packet will rebroadcast it until it has reached the destination node or it has a route to the destination. Such a node then replies with an Route Reply (RREP) packet, which is routed back to the source [96].

A number of strategies have been developed to establish asymmetric routes but these normally rely on establishing connectivity at full transmit power and then decreasing the node transmit power once the optimal routes have been determined. Route discovery information is distributed separately for each direction (forward and reverse path). The previously proposed methods maintain the connectivity relies on bidirectional routing with heavy control traffic. This result in complex routing operations is accompanied with redundant routing overheads. The algorithm provided here overcomes the reciprocity assumption in order to establish the most appropriate routes through a network. Moreover, it can operate over multi-hop networks where source and destination are not indirect contact when transmitting at full transmit power [96-100]. Tailoring the transmit power and routes through the network to fit in with desired constraints (e.g. hop count, battery life, interference) calls for cross layer design of the routing algorithm.

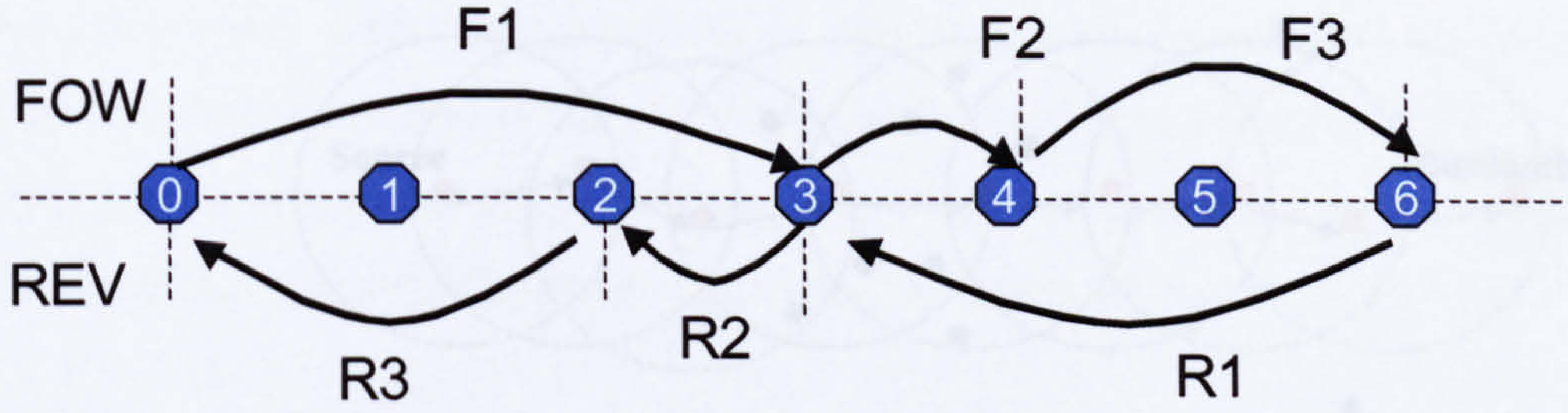


Figure 25 Asymmetric Routing Configurations[96-100]

If we consider the topology shown in Figure 25 and we assume that the forward path is a route we wish to determine, then the arrows depict the maximum link lengths that can be achieved, during this in a determinate time interval

Conventional route selection constraint often relates to shortest path (either physical path or minimum number of hops), minimum energy consumption, or selecting nodes with longest battery life. The latter two require cross-layer knowledge to be used, especially as they will often adjust the transmit power [96-100].

Figure 26 shows a scenario, where a path in general has a greater number of hops, resulting in longer packet delays. If variable transmit power is used, the transmit power can be varied to disturb a set number of nodes. In this way high transmit powers can be used in not congested parts of the network, resulting in fewer hops, with lower transmit powers used on parts of the route that are in close proximity to other networks. This provides more of a balance between disturbance and packet delay.

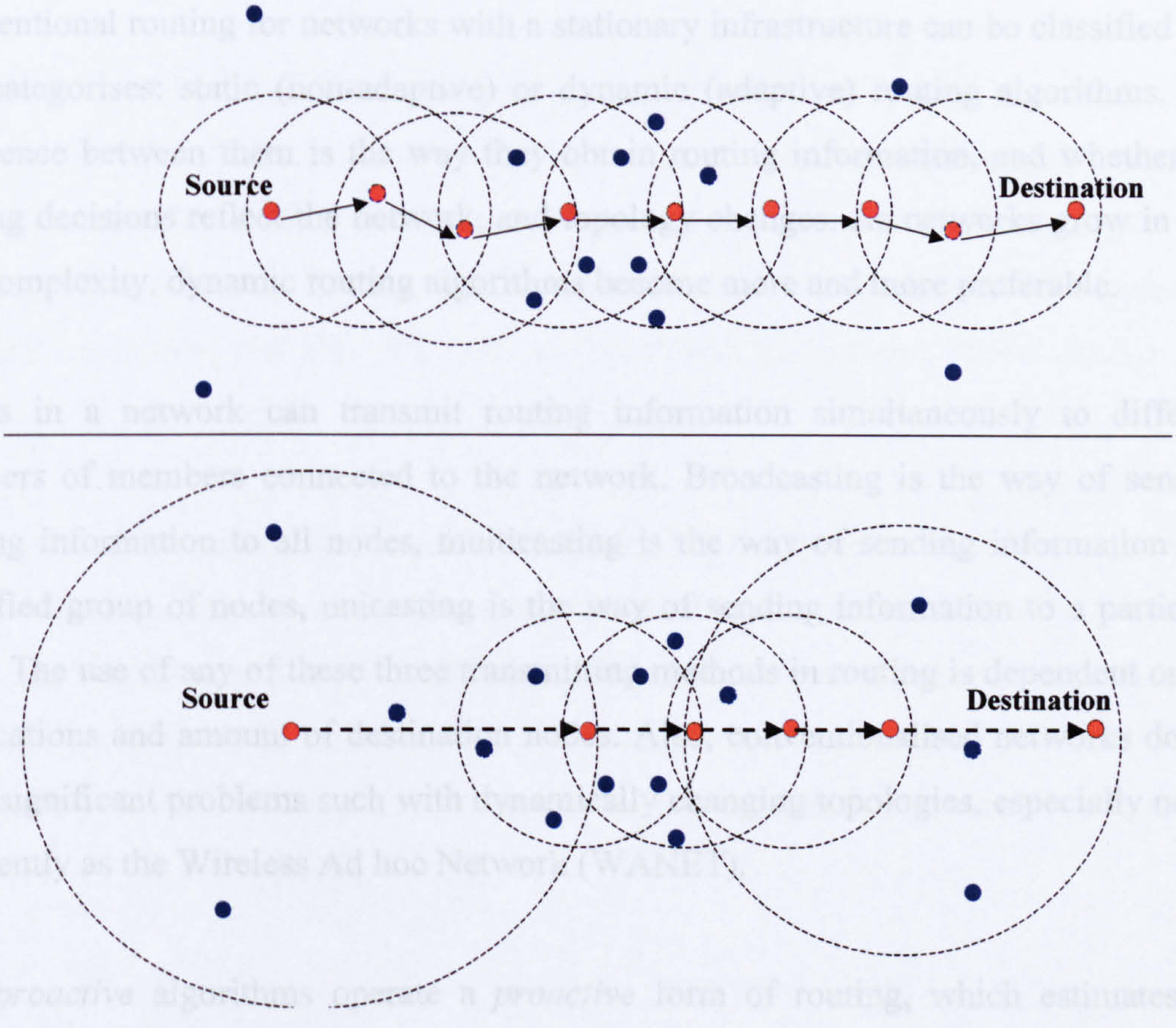


Figure 26 Scenario illustrating benefits of minimum impact routes

3.5 Conclusion

The routing related issues reviewed in this chapter start with a brief introduction of routing environment in wireless networks, which are related with four major issues: interference, transmit power, link length, and node density. These factors are essential for developing efficient routing strategies in WANETs, as well as enhancing capacity, to improve system performance.

Different routing algorithms introduced later in this chapter provide some important concepts that form part of later design and modelling of Minimum Impact Routing (MIR), which is the foundation of this research. The routing algorithm review includes fifteen classical algorithms, which are classified into three categories: table-driven, source-initiate/on-demand and hybrid routing.

Conventional routing for networks with a stationary infrastructure can be classified into two categorises: static (non-adaptive) or dynamic (adaptive) routing algorithms. The difference between them is the way they obtain routing information, and whether the routing decisions reflect the network, and topology changes. As networks grow in size and complexity, dynamic routing algorithms become more and more preferable.

Nodes in a network can transmit routing information simultaneously to different numbers of members connected to the network. Broadcasting is the way of sending routing information to all nodes, multicasting is the way of sending information to a specified group of nodes, unicasting is the way of sending information to a particular node. The use of any of these three transmitting methods in routing is dependent on the applications and amount of destination nodes. Also, conventionalised networks do not have significant problems such with dynamically changing topologies, especially not as frequently as the Wireless Ad hoc Network (WANET).

The *proactive* algorithms operate a *proactive* form of routing, which estimates the shortest-path between source and destination pairs. In general, it provides advantages such as short initial route discovery delays, and is relatively simple to implement. However, it may generate a significant amount of control traffic, which wastes network resources in maintaining routes that may never be used.

Reactive algorithms operate a *reactive* form of routing, which establishes and maintains a route only when it is in demand of a source node. It prevents constant maintenance of routes to all destination nodes, but only initiates a route discovery procedure in response to a route request. Further more it reduces the amount of control traffic generated by route discovery and maintenance, yet it has longer route discovery delay since routes have not been previously recorded.

Hierarchical routing algorithms provide a flexible solution using hierarchical routing, which constructs routing clusters/zones that contains all recognised wireless nodes. Proactive routing is used inside each cluster, and reactive routing is used for cross cluster route discovery and maintenance. Therefore, it reduces the control traffic and intra-zone route discovery delay; network loads dominated by control traffic are thus prevented. However it is complicated to implement such an algorithm.

The simplicity of proactive form of routing can be incorporated with the efficiency of the reactive form of routing, and can be incorporated into the development MIR. The polling strategy in CGSR can be used to solve the problem such as interference between adjacent nodes. The next chapter will introduce the simulation tool that used in network modelling, and this will help to understand the network model design used later.

The *asymmetric* routing algorithms distribute routing information separately on forward (source to destination) and reverse (destination to source) paths. This is due to the different transmitting power utilization or battery reserve remaining in each node. Therefore in real world applications, it is vitally important to take into account the asymmetric feature of wireless ad hoc networks in system, protocol, or standards design.

Chapter 4 Simulation and Verification Methodology

Contents

4.1 Introduction

Mixtures of simulative and analytical methods have been employed to investigate, analyse, and evaluate the behaviour and performance of WANET under different scenarios and configurations. The cross layer routing metric design and protocol development for this research conducted uses software, namely OPNET and MATLAB, for simulation and modelling, and the supporting theoretical analysis like Queueing Theory, Traffic Engineering Theory, Probability, and Stochastic Process Analysis. The computer-based test-bed are time and economically cost-efficient for abstractive and flexible comparison of different routing algorithms and protocols, hence above methods are the primary tools in this research.

Software *simulation* has certain limitations, such as approximated assumptions due to imperfect implementation or programming of the system details, and time consuming programming due to the complex structure as well as debugging process. On the other hand *analytical* models under these assumptions can provides table, graph, or diagrams to assist performance evaluation, and as a way of comparing with simulations in order to validate results.

In order to observe, analyse, and even predict the performance of WANET behaviour, we must first, analysis the behaviour or performance of WANET system use Queueing Theory. Then simulate WANET operations use OPNET simulation and modelling environment. This is followed by validation, evaluation, and analysis, using MATLAB to plot the collected raw data from OPNET simulation.

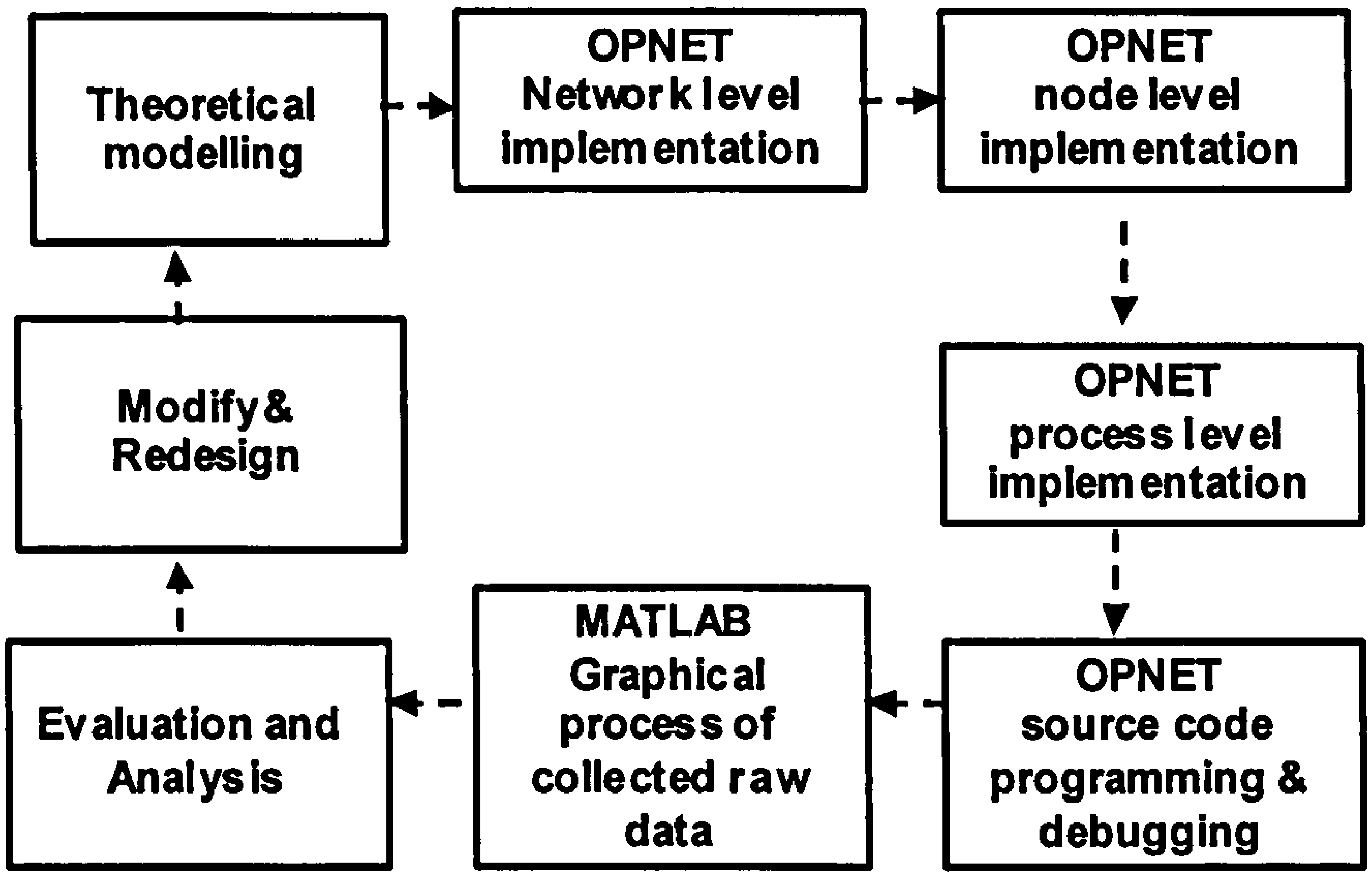


Figure 27 Design and Modelling Cycle

OPNET is a complicated network simulator, and a brief introduction of its modelling structure is necessary for understanding the later evaluation, and validate routing algorithm model. The design and modelling cycle apply to this research is shown in Figure 27.

4.2 Mathematical and Operational Modelling

We study the phenomena of the interference impact on WANET system capacity using analytical theory derived using Queueing Theory and stochastic analysis. The operational modelling simulates, using OPNET, the interaction between transmitting nodes that use different routing schemes.

4.2.1 Mathematical Analysis

Any system in which arrivals place demands upon a finite-capacity resource may be termed a Queueing System [31, 32, 101]. The Queueing System theory can describe the complex behaviour of a single terminal (such as a transmitting node, or a channel), or a communication network that contains multiple servers (e.g. WANET, or a source and destination pair with multiple intermediate relaying nodes).

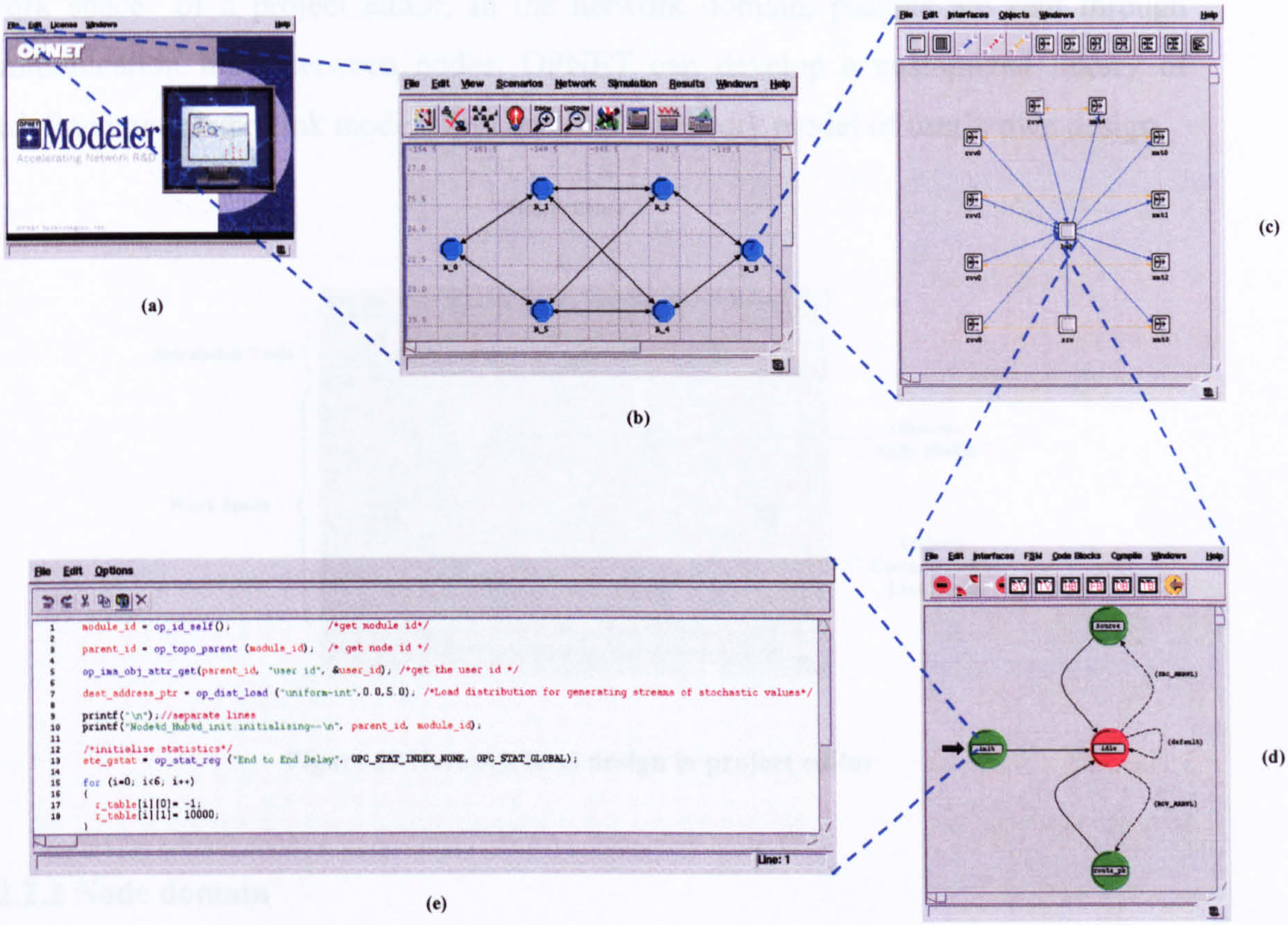
4.2.2 Operational Simulation using OPNET

OPNET is an event based, interrupt driven simulator, originally developed at Massachusetts Institute of Technology (MIT), and introduced in 1987, for network design and protocol simulation. OPNET Modeler (short for OPNET) contains different integrated tools that support modelling and simulation of cable-based/wired systems, satellite systems, mobile and fixed wireless systems [102].

OPNET simulation is event-driven, and it constructs packets to transfer data rather than simulating bit streams. It uses "event time stamp", which associates each event that simulated with a specific simulation time, therefore the simultaneously occurring events at different nodes can be accurately modelled and recorded, and the simulation executed on a conventional serial-based computer simulating those events of nodes operating in a parallel manner. OPNET is packaged with a model library, which contains standard network models, and protocols (e.g. ATM, TCP/IP and RIP etc). OPNET using a graphical interface/window for most stages of a simulation, and it is powerful, because most functions of systems can be represented in a graphical form.

OPNET simulates systems in a hierarchical fashion, and generally, a model is specified in three levels: network level, node level and process level. Each level defines one modelling domain, and it is different from most modelling frameworks, which specify all aspects of a system by using a single paradigm. OPNET specifies features of each level using a hierarchical set of editors, namely, Project Editor for the network level, Node Editor for node level, and Process Editor for process level.

Figure 28 illustrates an example of a cable-based network model, in which a model is constructed in four hierarchical editors. Figure 28 (b) is the project editor for network level/domain design; (c) is the node editor for node level design; (d) is the process editor for process level design; (e) is a specification panel, which defines logical states in a processor block of a node, from a source code level.



```
module_id = op_id_self();                      /*get module id*/

parent_id = op_topo_parent (module_id);  /* get node id */

op_ima_obj_attr_get(parent_id, "user id", &user_id); /*get the user id */

dest_address_ptr = op_dist_load ("uniform-int",0.0,5.0); /*Load distribution for generating streams of stochastic values*/

printf("\n");//separate lines
printf("Node%d_Hub%d_init initialising--\n", parent_id, module_id);

/*initialise statistics*/
ete_gstat = op_stat_reg ("End to End Delay", OPC_STAT_INDEX_NONE, OPC_STAT_GLOBAL);

for (i=0; i<6; i++)
{
  r_table[i][0]= -1;
  r_table[i][1]= 10000;
}
```

Figure 28 OPNET Model Hierarchical Structure [102]

(a) OPNET Modeler starting window (b) Project Editor for Network Level

(b) Node Editor for Node Level (d) Process Editor for Processor level

(e) State Specification in Proto-C for source-code level.

4.2.2.1 Network domain

The highest layer of the hierarchy is the *network domain*, which specifies high-level devices (i.e. nodes and communication links) and the topology of a *network model* uses a *Project Editor*. A network may contain any number of nodes, which can be placed on a geographical area, and different network layers may be configured through multiple layer nesting. It can also specify the orbit of a satellite or a mobile node. The same type of Hexagon routers symbol represent the wireless nodes later developed in the MIR software models.

Figure 29 illustrates an example of 6-node-network with cable link connections in the “work space” of a project editor. In the network domain, packets are sent through communication links between nodes. OPNET can develop a customized library of predefined nodes and link models for construct a network model of user’s own design.

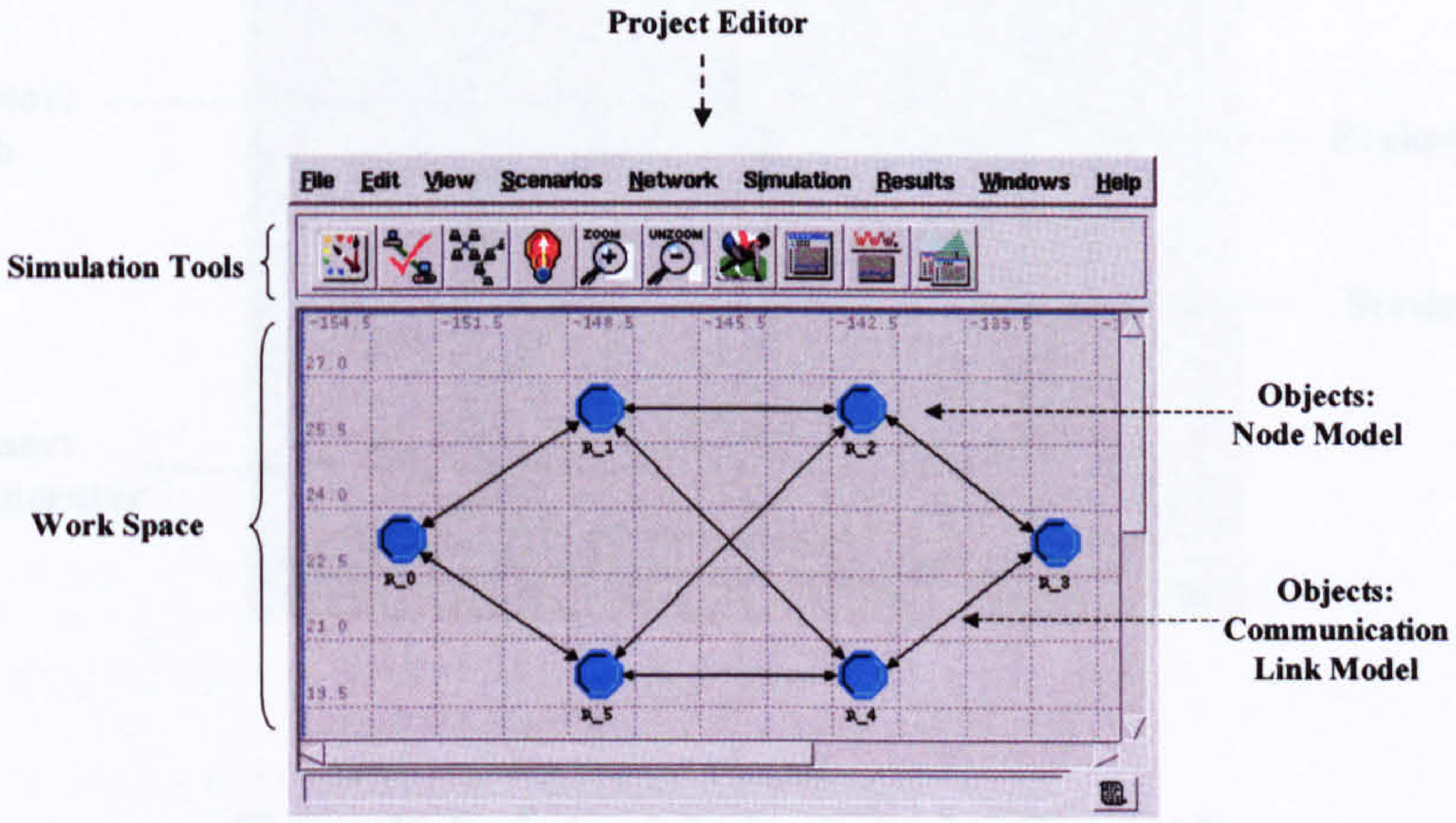


Figure 29 Network level design in project editor

4.2.2.2 Node domain

The second layer of the hierarchy is the *node domain*, which specifies the functional elements of each *node model* using the *Node Editor*. The node model represents a communication node of a network model, which is defined in the network domain, and the functional elements of the node is a *processor block*, which is capable of performing a specific task in the node, (e.g. transmitter, receiver, hub and packet generator).

The node model can be used to represent any network equipment such as routers, workstations, terminals, servers, switches and satellites etc. Packets are transmitted through *packet streams* connecting these processor blocks. *Statistic wires* are used to transfer discrete statistics between processor blocks. Packet streams and statistic wires are connections and logical associations to allow information flow between processor blocks.

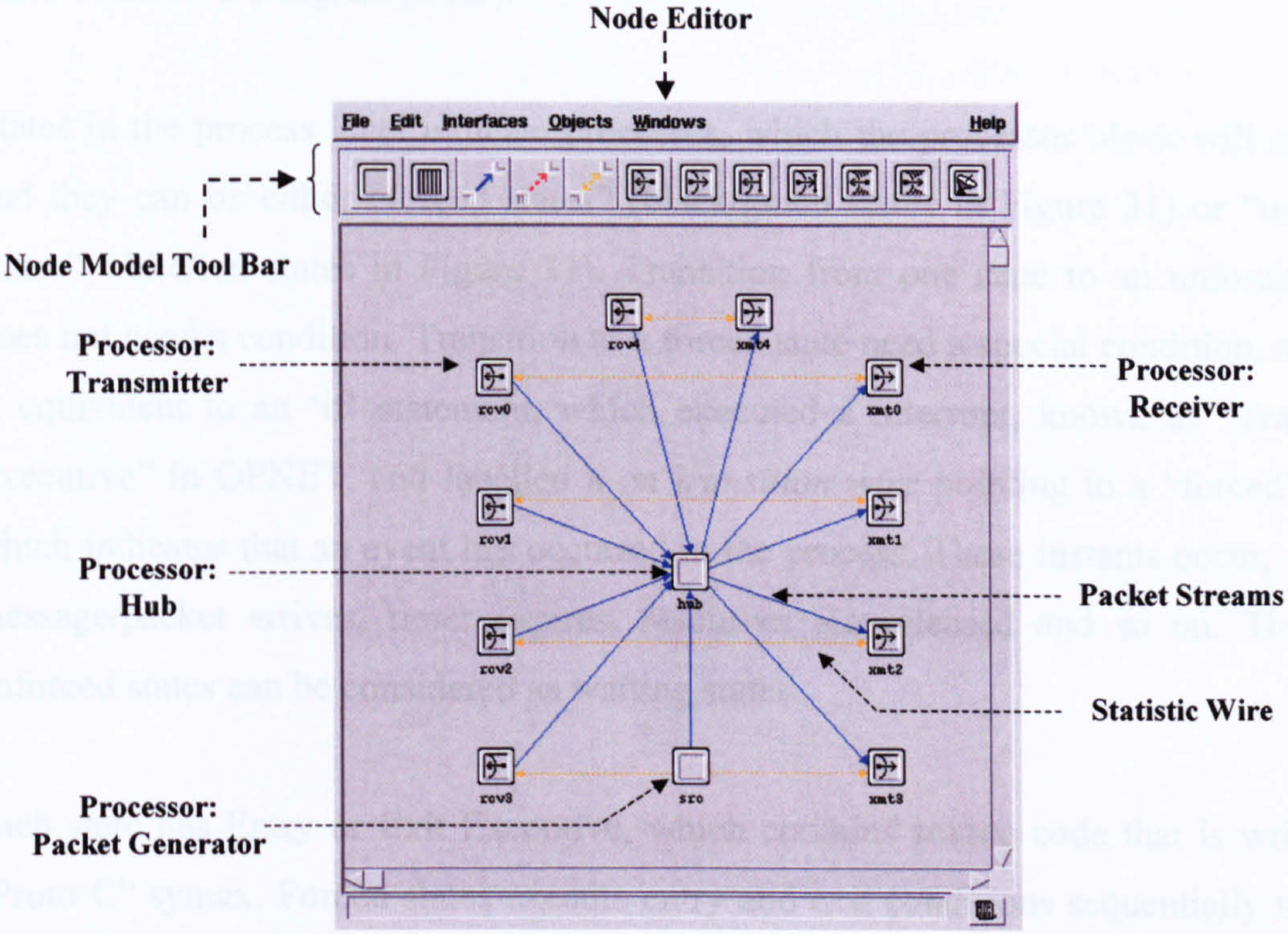


Figure 30 Node level design in node editor [102]

Figure 30 illustrates a node level design for communication node (R1) shown in Figure 29 above. This node model design consists of processor blocks for executing the function of: hub, transmitter, receiver (5 pairs) and packet generator of the node model.

4.2.2.3 Process domain

The lowest layer in the hierarchy is the *process domain*, which defines the *process model* of a processor block using the *Process Editor*. The process model describes, in the process level, the operating process or state of a processor block, which is defined in node level previously. Process level is where the protocol mechanisms or routing algorithm are designed and developed, and it takes most of the modelling time to describe states when developing a new scheme or algorithm.

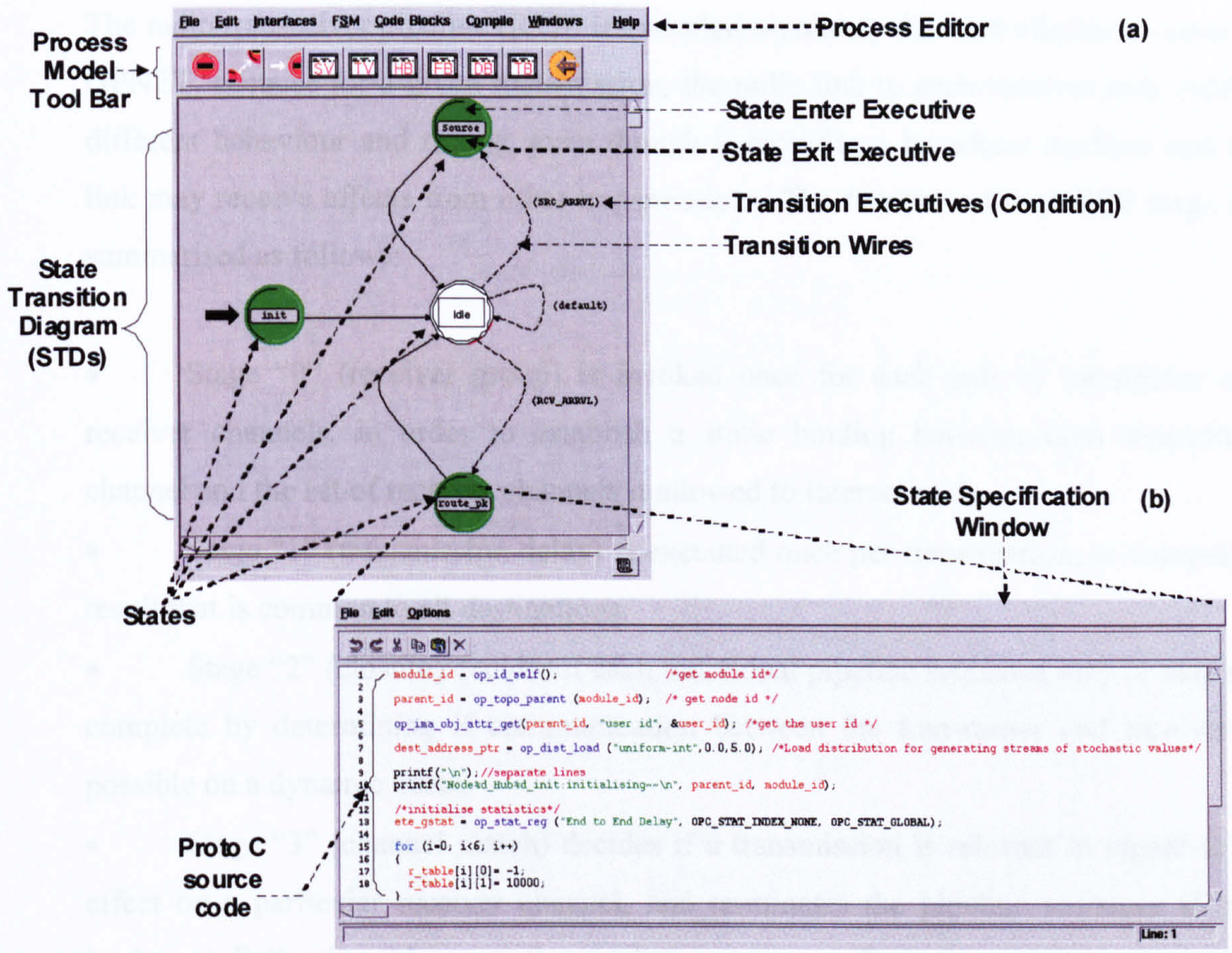
Process models are designed using a graphical extension of the "C" computer language, called "Proto C" [102], which consist of states, and it is similar to those states used in Finite State Machine (FSM). At any particular time, only one state is active, and a simulation may execute at this state and then transfer control to another state later. A series of states are connected together by transition wires, organized into a graphical State-Transition-Diagram (STD).

States in the process level indicate processes, which the processor block will execute, and they can be either "forced states" (black/green states in Figure 31) or "unforced states"(white/red states in Figure 31). Transition from one state to an unforced state does not need a condition. Transition to a forced state need a special condition, and this is equivalent to an 'if' statement, which executed a interrupt, known as "Transition Executive" in OPNET, and labelled a on *transition wire* pointing to a "forced" state, which indicates that an event has occurred in the process. These instants occur, when a message/packet arrives, timer expires, resources are released and so on. Therefore unforced states can be considered as waiting states.

Each state has Entry or Exit Executive, which contains source code that is written in "Proto C" syntax. Forced states execute entry and exit conditions sequentially with no pause between them; the unforced states execute with a pause between them, with the exit condition being executed on the next interrupt.

The base facilities in "Proto C" use conventional "C", and in addition, "Proto C" has a library of high-level commands known as "Kernel Procedures (KPs), which are equivalent to "C" functions.

Each process model in a node operates in parallel with all the other models, and this allows OPNET to simulate a real network, which each processor in a node operates independently. The parallel event simulation is achieved by time stamping every event with an instantaneous simulation time.



```
module_id = op_id_self();              /*get module id*/

parent_id = op_topo_parent (module_id);  /* get node id */

op_ima_obj_attr_get(parent_id, "user id", &user_id); /*get the user id */

dest_address_ptr = op_dist_load ("uniform-int",0.0,5.0); /*Load distribution for generating streams of stochastic values*/

printf("\n");//separate lines
printf("Node%d_Hub%d_init:initialising~~\n", parent_id, module_id);

/*initialise statistics*/
ete_gstat = op_stat_reg ("End to End Delay", OPC_STAT_INDEX_NONE, OPC_STAT_GLOBAL);

for (i=0; i<6; i++)
{
  r_table[i][0]= -1;
  r_table[i][1]= 10000;
}
```

Figure 31 Processor domain

(a) Processor Editor (b) State Specification Window[102]

Figure 31 (a) illustrates the process level design, which is a STD of the processor block "hub" that defined in Figure 30 above. There are four states in the STD: initial state, idle state, route arrived packet (Route_pk) state and route local generated packet (source) state. Figure 31 (b) illustrates the entry executive defined by "Proto C" source code written in a state specification window.

4.2.2.4 Radio Transceiver Pipeline

Radio links do not exist as physical object in an OPNET simulation, therefore the radio transmissions are simulated using radio pipeline stages, which support a particular radio transmission, and are associated with the radio transmitter and receiver.

The radio transceiver pipeline (RTP) is executed separately for each eligible receiver in OPNET, because for a given transmission, the radio link to each receiver may exhibit

The radio transceiver pipeline (RTP) is executed separately for each eligible receiver in OPNET, because for a given transmission, the radio link to each receiver may exhibit different behaviour and timing, even though it provides a broadcast medium and the link may receive affects from other transmissions. The function of each RTP stage are summarised as follows:

- Stage "0" (receiver group) is invoked once for each pair of transmitter and receiver channels, in order to establish a static binding between each transmitter channel and the set of receiver channels it allowed to interact with.
- Stage "1" (transmission delay) is executed once per transmission, to compute a result that is common to all destinations.
- Stage "2" (closure) decides if each individual pipeline sequence may or may not complete by determining if communication between the transmitter and receiver is possible on a dynamic basis.
- Stage "3" (channel match) decides if a transmission is relevant in regard to its effect on a particular receiver channel, and terminates the pipeline sequence if it is irrelevant. Following this stage, the pipeline sequence will continue or stop progress, if the pipeline sequence progress, the packet transmitting is approved as a **valid packet,** and then the process continuous.
- Stage "4" (Tx antenna gain) and stage "6" (Rx antenna gain) will determine the antenna gain associated with transmitter and receiver.
- Stage "5" (propagation delay) calculates the time required for the packet's signal to travel from the transmitter to the receiver.
- Stage "7" (received power) computes the received power of the arriving packets signal in watts.
- Stage "8" (background noise) represents the effect of all noise sources.
- Stage "9" (interference noise), stage "10" (signal to noise ratio), stage "11" (bit error rate), stage "12" (error allocation) evaluate a link's performance in response to changes in the signal condition.
- Stage "13" (error correction) determines whether or not the arriving packet can be accepted and forwarded via the channel's corresponding output stream to one of the receiver's neighbours.

Figure 32 illustrates an example of the RTP, which consists of these fourteen stages [102]. OPNET executes these computational states to simulate radio transmission between transmitter and receiver.

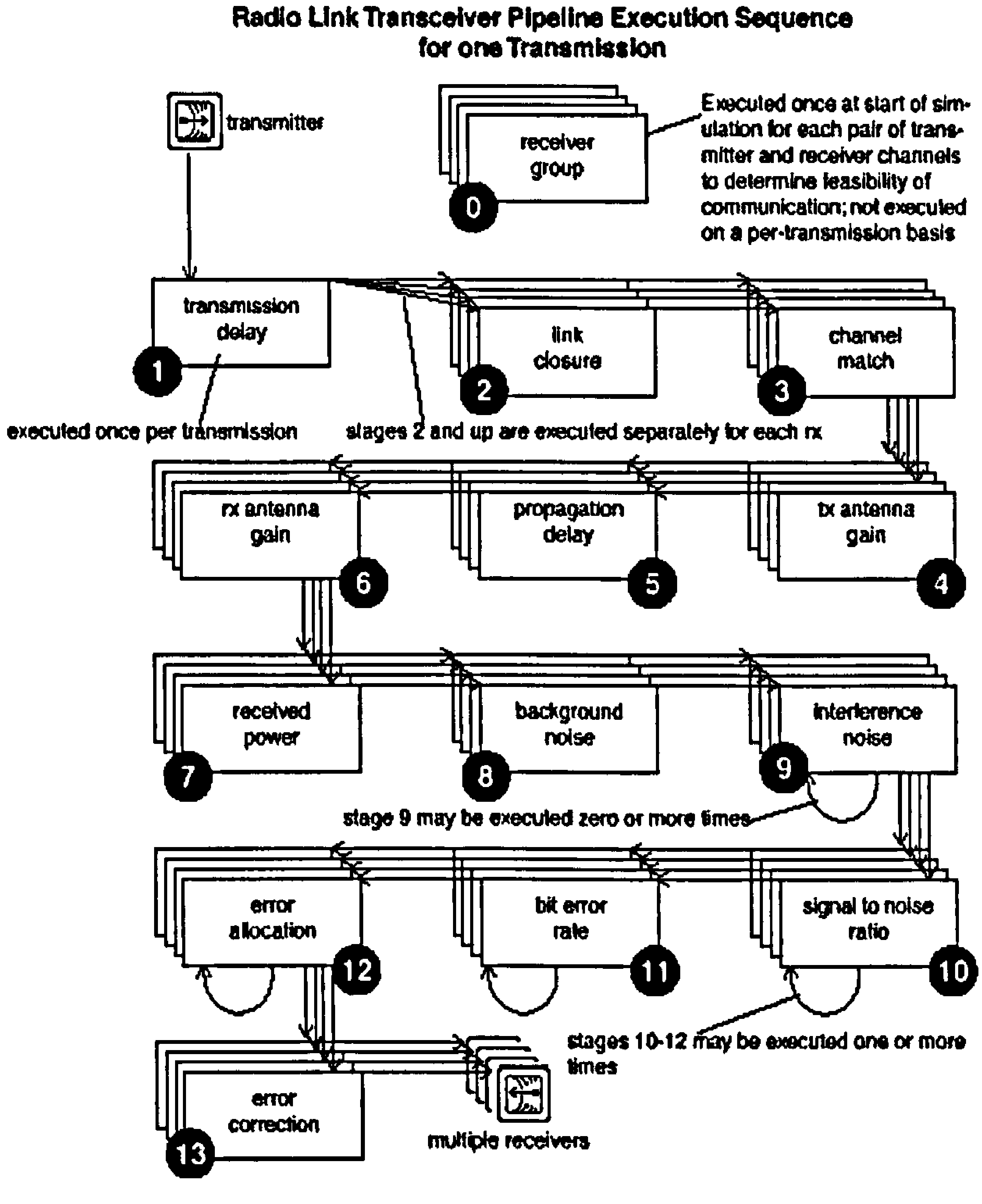


Figure 32 Radio Transceiver Pipeline in OPNET [102]

4.3 Validation of Simulation Results use MATLAB

The analysis, evaluation, and validation process for the mathematical and operational model are carried out by use OPNET and MATLAB. The 'global variables' in the OPNET simulation model enables the statistical data to be collected during the simulation, and a text file, collects the raw data during the simulation, that reflects the performance of the WANET model, which is made available to the end of the simulation process. These raw data files then will be processed using MATLAB to produce analytical plots, rather depends only on the data analysis tools that embedded in OPNET.

Considerable effort has been devoted to validate the simulation and analytical models and results presented in this thesis. The reproductions of the classical routing algorithm e.g. Shortest Path (SP), is compared with novel Minimum Impact Routing (MIR), under different network environments. A wireless network with regulated and random topology that can be configuring into multiple scenarios is used as a simulation test-bed for system performance analysis.

The theoretical analysis of the WANET system is simulated use the model mentioned above, the statistical and analytical results that validate these theories are presented throughout chapter 6 and 7. Various of measure, such as disturbed nodes, hop counts, delays, generated traffic, and throughputs, are made for collect raw data in order to produce validation results.

4.4 Conclusion

This chapter overviews the *analytical and statistical procedures* and *operational simulation* methodology for analysis system performance, statistical results and evaluate the performance of WANET under different scenarios use MIR routing strategies.

The advantage of the OPNET approach has been discussed when the problem of modelling considers both system's behaviour and its structure. For example, communication protocols and real-time operating systems are different types of entities from communication links or packet buffers.

OPNET modelling is based on three paradigms, which specifically target the distinct levels identified in a communications network. The single-paradigm modelling must be stretched to develop adequate models.

OPNET provides a useful tool and powerful support in simulation and analysis of network protocols. However it is complicated to learn and takes significant amount of time before a complex algorithm can be effectively implemented. Therefore a detailed understanding of behaviour of the standard OPNET model and the radio transceiver pipeline is essential for modelling a wireless system.

Verification of results is provided where possible by comparing performance using a mixture of analytical and simulation technique Statistical data that collected throughout the simulation use OPNET model. These data, which representing various of system performance measures, are processed use MATLAB to produce sensible analytical plots and graphs that presented in chapter 6 and 7.

Chapter 5 Minimum Impact Routing Preliminary

Contents

5.1 Introduction

In order to understand how a routing mechanism may affect the system performance, an analysis of system behaviour must be carried out prior to the routing algorithm design. This chapter identifies, in section 5.2, the pivotal trade-offs used in evaluating the WANET system performance, namely the *system supported user population*, system *traffic handling capability*, and *system scalability*. In section 5.3, various solutions for these trade-offs are proposed in order to resolve the resulting challenges. These techniques have become the inspiration for the design of the Minimum Impact Routing (MIR) scheme, with its premier objective dedicated to resolving the system scalability problem in WANET by minimising the interference impact and maximising the spatial reuse in the system.

5.2 Cross-Layer Challenges

WANET was initially known as a Packet Radio Network, which is a special form of packet switched wireless communication network. The system capacity analysis and mathematical modelling of the system behaviour quite often forces researchers to make simplified or idealized assumptions for their analytical results. However, the analysis in this chapter will investigate system behaviour, under the some strong assumptions, from a cross-layer aspect. There is strong influence from the *physical* and *multiple access control* layer on the routing activity, which relates more to the n*etworking* layer. Some of the well-known fundamental principles from *Queueing Theory* are mentioned to assist the analysis.

WANET can be considered a *Queueing System* where each terminal is a potential server that processes a flow of traffic. In such a system, the system performance is generally measured by *throughput*, *delay*, *system supported user population,* or *system utilization* performance, as shown below:

- **Throughput:** Number of Tasks (N) = Average Arrival Rate (λ) × Delay (T)

  $N=\lambda T$ Or

  Successful Throughput λ(s) = Offered Load λ(o) × Probability of k Success (P_k)

  $\lambda(s)= \lambda(o)\ P_k$
- **Delay:** System Response Time (T) = Waiting Time in Queue (w) + Service Time (x)

  $T = w + x = N/\lambda$ Or

  $T = 1/(\mu - \lambda)$ (when N=ρ/(1-ρ) in M/M/1 Queue System)
- **User Population:** User Population supported by the system. E.g. in cellular system is: ((Total Available Bandwidth (B_t) /Channel Bandwidth (B_c))/Cluster Size (C_n))*Number of Cells (N) in the system.

  User population = $((B_t/B_c)/ C_n)\ N$
- **System Utilization:** Total busy period of the system can be represented as follow,

  Utilization (ρ) = Average Arrival Rate (λ) * Service Time (x) = Average arrival rate (λ) / Average Service Rate (μ).

  $\rho= \lambda x =\lambda/\mu$
- **Relaying Burden:** Number of Accumulated Tasks (N_i) =Number of Simultaneous (Transmitting Node (n_t)*Number of Neighbouring Node (n_e) increases exponentially to the power of Number of Hops (H).)

$$N_l = n_t \, n_e^{H}$$

- **Reuse Ratio:** Reuse Ratio (S)= per path Occupied Resources (N_l)/per hop Occupied Resource (n_l)

$$S = N_l / n_l$$

In WANET, the choice of a routing mechanism that can support more users is one of the first well-explored challenges. However, none of the proposed routing protocols can support a large user population as conveniently as cellular networks do today. In other words, the scalability of the WANET system is not as good as cellular systems. This measure of the user population that can be supported by a system leads to the first argument in terms of *single-hop* routing versus *multihop* routing.

5.2.1 User Population: Single Hop versus Multihop Routing

WANET terminals can communicate with any other member of the same system by using a one-hop link with high transmit power, or multihop links that use lower transmit power. If simultaneously transmitting nodes in both scenarios can reach the same per node throughput, then the overall system/network aggregated throughputs will be higher in a system with more simultaneous transmitting points, disregarding the level of service, relaying burden, and avalanche effects of flooding in multihop systems.

*The **first trade-off** concerns whether the system performs better using single-hop transmission for routing without relaying, in which case packets experience shorter delays but more nodes are interfered with instantaneously; or instead use multihop transmission for routing with more relaying packets, which allows more simultaneous transmissions.*

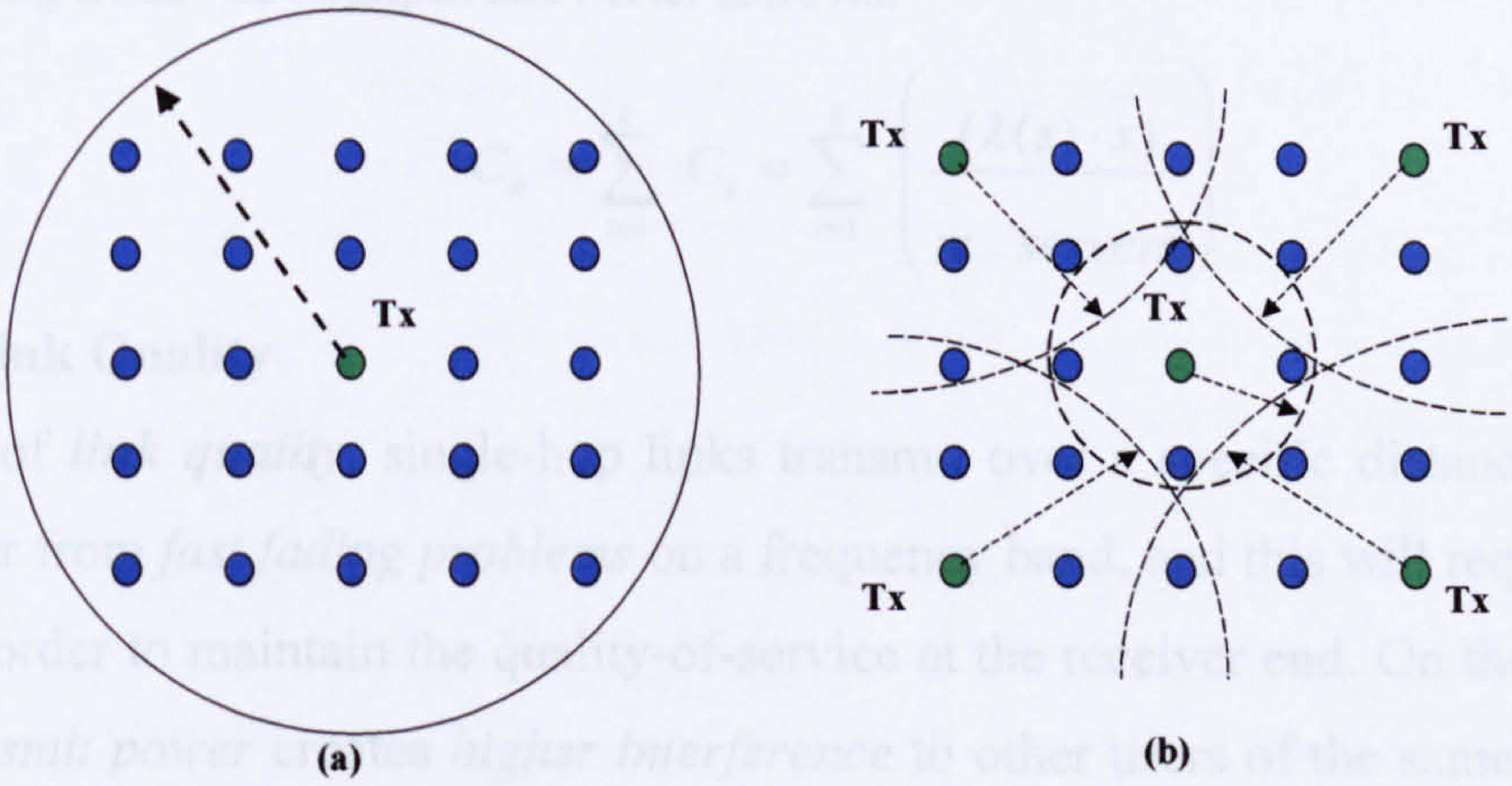


Figure 33 Networks with Different Link Length

(a) Single Hop System (b) Multihop System

5.2.1.1 System Throughput

In terms of ***system throughput***, consider two identical systems that have perfect scheduling, along with the highly abstractive assumption of system throughput that is not influenced by increasing relaying burdens. The system that has a higher number of simultaneous transmissions, could support a larger user population, which is a desirable feature [101]. Figure 33 (a) shows an example of a single-hop network, and (b) is the multihop case.

If there is only one simultaneous transmitting node, corresponding to *1 server* in a queueing system, operating at any moment in the system, then the aggregated system throughput, denoted by C_n in equation 5.1, equals the average throughput of an individual node, denoted by C_i, in which $\lambda(s)$ is successful throughputs of the offered traffic $\lambda(o)$, and $\bar{x}$ is the average service/delay time or system response time without queueing delay.

$$C_n = \sum_{i=1}^{k} C_i = \frac{(\lambda(s) \cdot \bar{x})}{1 \ \ server} \tag{5.1}$$

In multihop systems, multiple simultaneous transmissions, "*m*", provide the system with an aggregated throughput C_n that is equal to the sum of all simultaneously transmitting node' throughput shown as follows:

$$C_n = \sum_{i=1}^{k} C_i = \sum_{i=1}^{k} \left(\frac{(\lambda(s) \cdot \bar{x})}{m \quad servers} \right) \tag{5.2}$$

5.2.1.2 Link Quality

In terms of ***link quality***, single-hop links transmit over a specific distance, the signal will suffer from *fast fading problems* on a frequency band, and this will require a higher SINR in order to maintain the quality-of-service at the receiver end. On the other hand, *high transmit power* creates *higher interference* to other users of the same system, and this will also increase the chance of a *transmission error* and introduce more retransmission traffic into an already dense traffic flow. In the multihop case, signals transmit over a shorter link length with less fading, and the lower transmit power creates less interference to others, thus resulting in higher SINR and a lower error rate that leads to a lower retransmission rate and consequently lower retransmission traffic[10].

If only the above considerations of the system throughput or link quality are a tenable base to make, one could conclude that multihop routing is more desirable than the single-hop routing mechanism, due to the fact that multihop transmission supports a larger user population operating simultaneously, assuming perfect scheduling. However, in practice, routing in a multihop system is conducted by relaying, which results in unavoidable relaying delays and extra traffic forwarding burdens that may decrease the goodput over successes and furthermore, a system reducing the throughput per node to an unacceptable level.

In terms of ***delay performance*** of a system, the drawback of multihop transmission is the multiple relaying delays, which is *H* times (indicating number of hops) longer than a single-hop system.

$$EndToEndDelay(x) = \sum_{i=0}^{H} x_i = PerHopDelay(x_i) \times NumberOfHopsPerRoute(H) \tag{5.3}$$

Table 4 shows a summary and comparison of routing using single-hop or multihop transmissions in the system, and the corresponding node and system throughput at a transient moment.

	Single-Hop Routing Characteristics	**Multi-Hop Routing Characteristics**
Advantages	● High Connectivity ● Low Intermediate Relay & Delays ● Low Relay & Control Overheads	● Lower Interference ● Lower Error Rate (BER , PER) ● Lower Retransmissions Rate ● Higher Spatial Reuse ● More Scalable ● Lower Access Request Intensity ● High Contention and Collision
Disadvantages	● High Interference ● High Error Rate (BER, PER) ● Higher Retransmissions Rate ● Low Spatial Reuse ● Less scalable ● High Access request intensity ● High contention and collision	● Weak Connectivity ● High Intermediate Relays & Delays ● Higher Relay & Control Overheads
Average Node throughput	$C_i = \lambda(s) \cdot \bar{x}$ *no relays*	$C_i = \frac{\lambda(s) \cdot \bar{x}}{m}$ *with relays*
Average System throughput	$C_n = C_i$ *no relays*	$C_n = \sum_{i=1}^{k} C_i$ *with relays*

Table 4 Comparisons of Single-hop and Multihop Routing

Apart from the user population consideration, the relaying burden and the avalanche effect due to the use of flooding relating to the routing algorithm leads to the second sets of arguments that are covered in next section.

5.2.2 Traffic Handling: Centralized versus Distributed Control

If the single-hop and multihop discussion is more related to the *physical* (PHY) layer issues such as SINR, error and retransmission rates, and interference, then there is a second trade-off between centralized and distributed control, which more relates to *Multiple Access Control* (MAC) layer issues, namely, the contention control, and overheads due to control and relaying burdens.

*This **second trade-off** leads us to the decision of whether to use centralized control, which provides a lower number of relays and shorter delays, hence lower relaying overhead, but on the other hand, has a large number of neighbouring node access contention or to use distributed control that has lower contention control overhead due to the use of localized routing information from a lower density of neighbouring nodes.*

In terms of ***system utilization and efficiency***, the system throughput consists of two portions of data, one part is the useful information reaching its preferred destination, and the other part is the overheads. Goodput is the application level throughput that measures the number of useful bits per unit time forwarded from a source to a destination, excluding overheads of the protocol, control, relaying, or retransmitting data packets.

$$Throughput = Goodput + Overheads \tag{5.4}$$

The channel or system throughput will consist of a smaller proportion of *goodput* if the overhead is high in the traffic flow.

5.2.2.1 Relaying

In terms of ***overhead control or reduction***, the *relaying burden* and the consequent *avalanche effect* due to flooding are critically important and give rise to exponentially growing control or protocol overheads. Overheads are extremely sensitive to the *neighbour node density* at each hop and the *route length* (measured in hops) of each routing task. *In WANET systems, the relaying burden is one of the major limitations for system or link throughput*[2, 11, 13-15, 17, 24, 27, 28, 43, 103, 104].

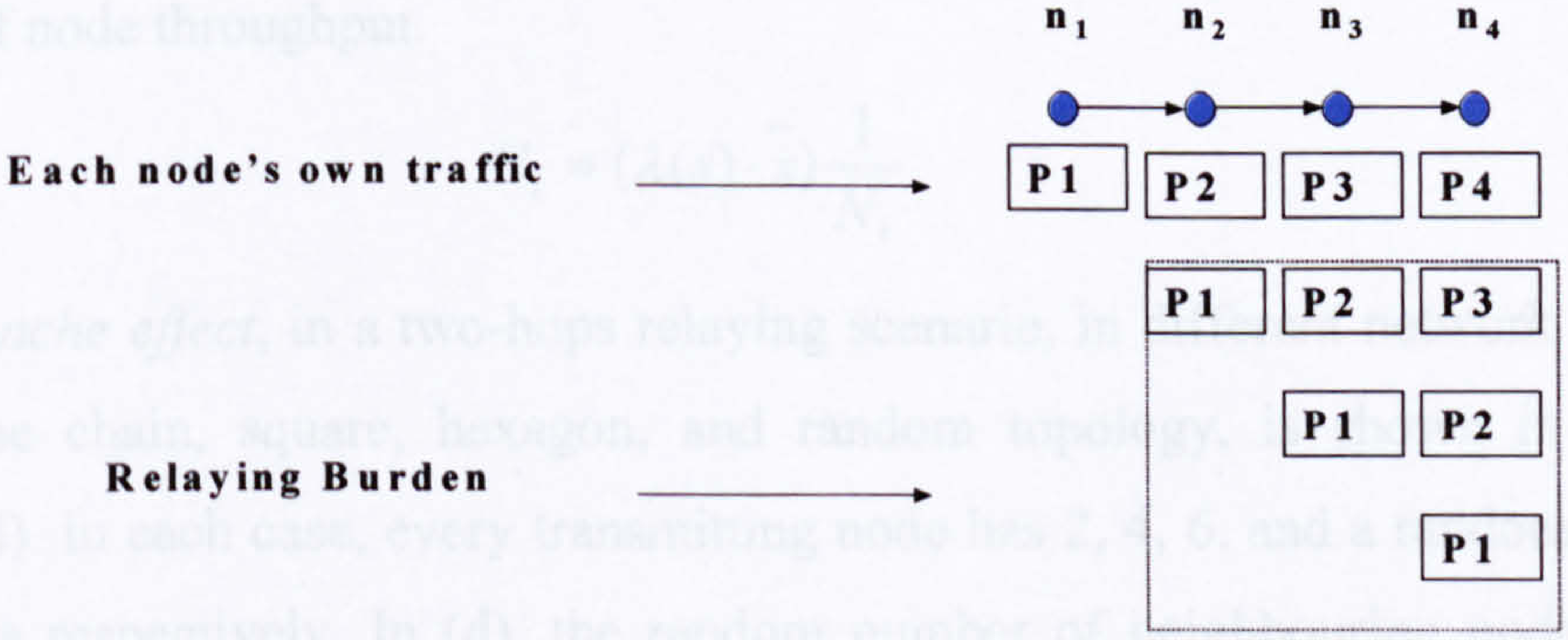


Figure 34 Relaying Burden along the path

The *relaying burden* is the unwanted forwarding tasks and traffic from other neighbouring nodes. Figure 34 demonstrates this by showing a simplified channel utilization case along with a three-hop relaying, in which the source node, n_1, generates a packet, *P1*, to be forwarded to its destination node n_4. Intermediate nodes n_2 and n_3 carry the relaying burden for n_1, at the same time transmitting traffic of their own, with the remaining 50% and 33% of their own channel or processing resource.

In the WANET environment, where all terminals share a common channel frequency, the *relaying* creates an extra burden for intermediate nodes. Furthermore, the flooding related routing scheme (mostly at route discovery stage) has an *avalanche effect* problem, which means it accumulates this relaying burden at each hop and can make packets duplicate exponentially out of control. Apart from limiting the residual life of a packet, and using a packet sequence number that refers to the packet source to reduce duplicates, maintaining an optimal local *neighbour node density* is the key to this problem.

Figure 35 shows the accumulated relaying burden denoted by N_l (later defined as end-to-end accumulated disturbed nodes) that is the accumulated per-hop interfered node, n_i, which consists of a cluster of neighbouring nodes, n_e, and a transmitting/relaying node, n_t, along the path.

$$N_l = n_t \cdot n_e{}^H = \sum_{i=0}^{H} \overline{n_i} \tag{5.5}$$

This N_l grows non-linearly in number that is in inverse proportion to the average node throughput, C_t, thus we can derive the following equation to measure relaying burdens in terms of node throughput.

$$C_t = (\lambda(s) \cdot \bar{x}) \frac{1}{N_l} \tag{5.6}$$

The *avalanche effect*, in a two-hops relaying scenario, in different network topologies, namely the chain, square, hexagon, and random topology, is shown in Figure 35 (a)(b)(c)(d). In each case, every transmitting node has 2, 4, 6, and a random number of neighbours respectively. In (d), the random number of neighbouring nodes surround each working node (Green dot) follows Poison Random Distribution (PRD).

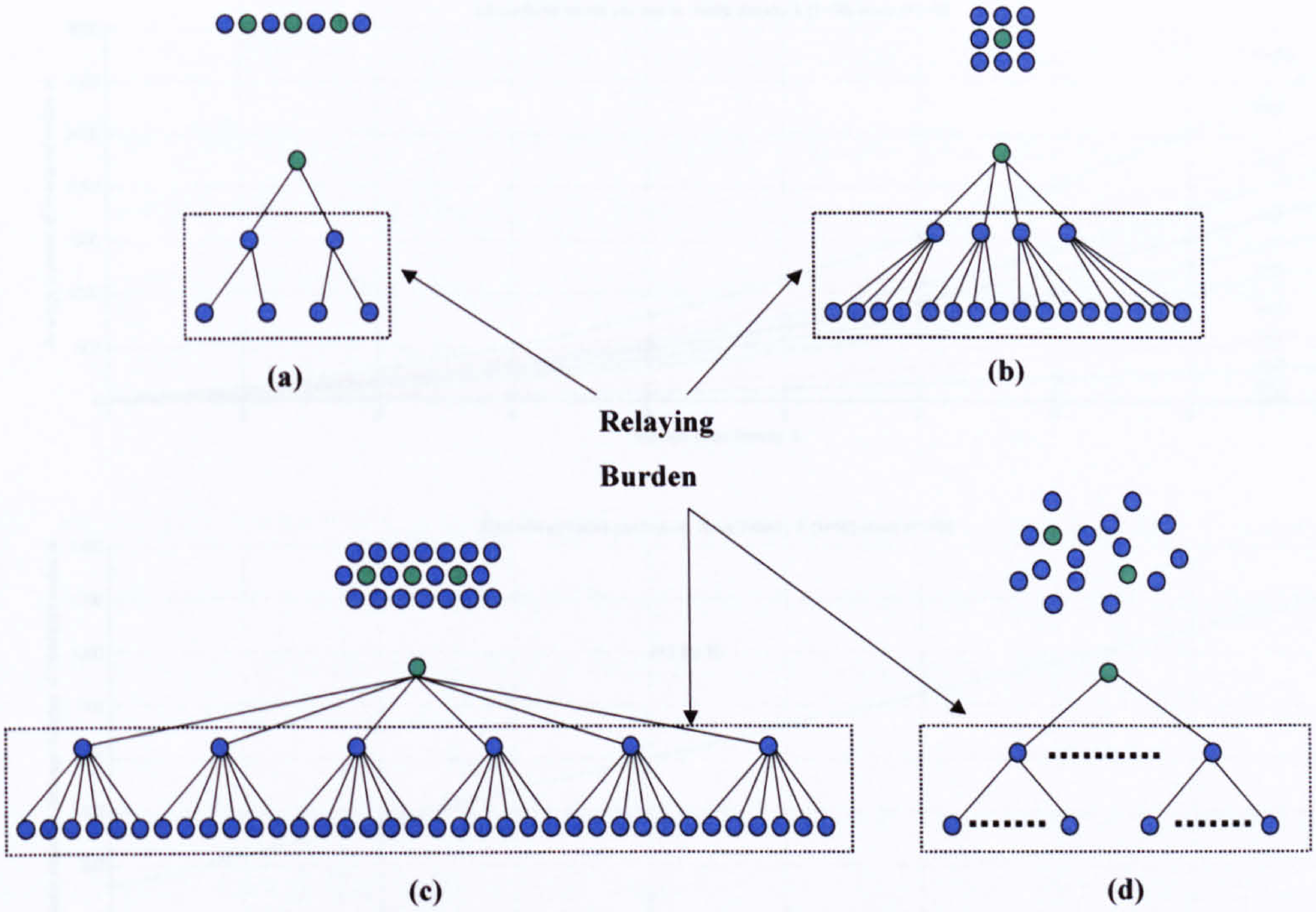
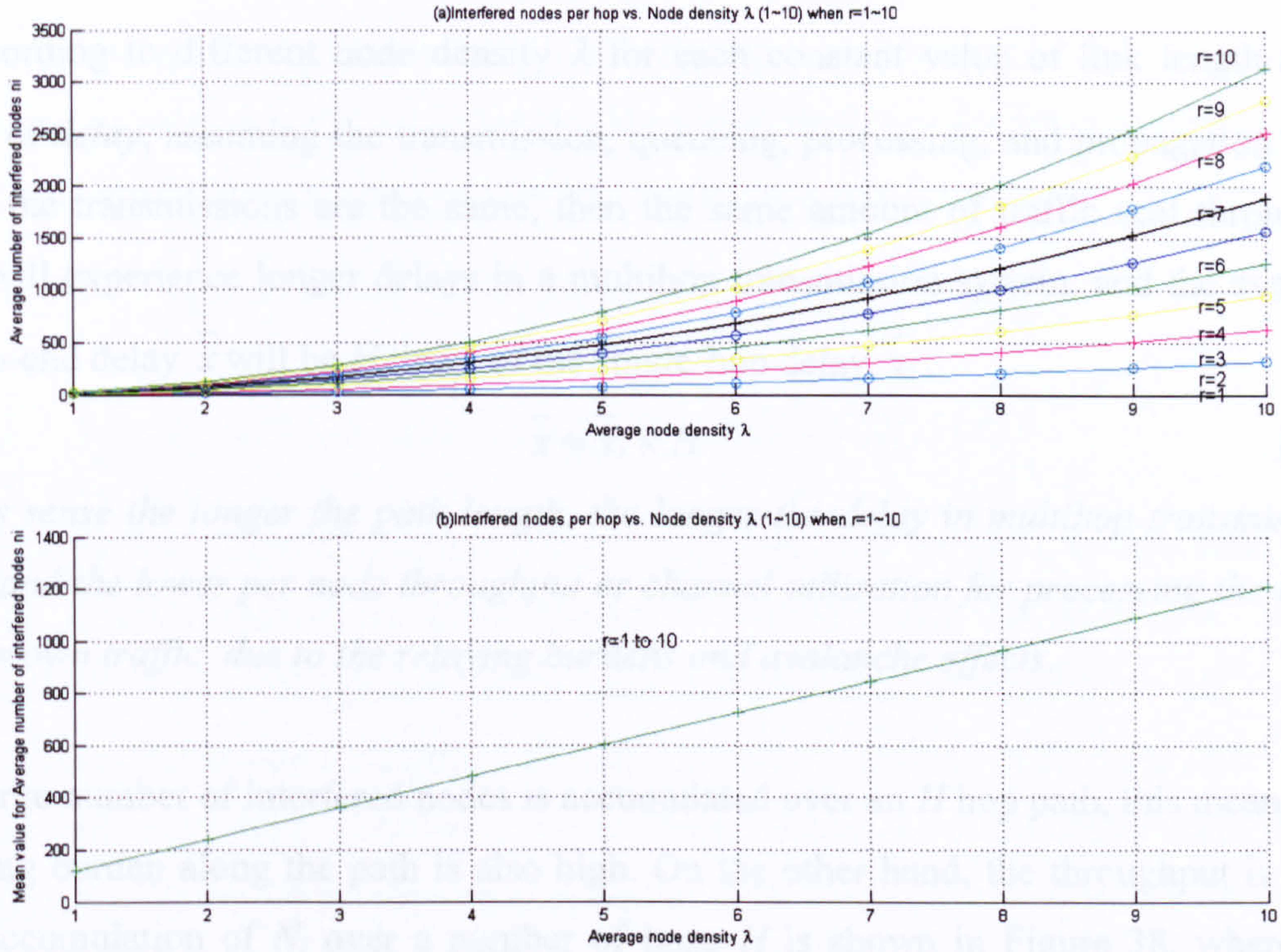


Figure 35 Avalanche Effect and Relaying Burden in different network topology

(a) Chain Topology (b) Square Topology (c) Hexagon Topology (d) Random Topology

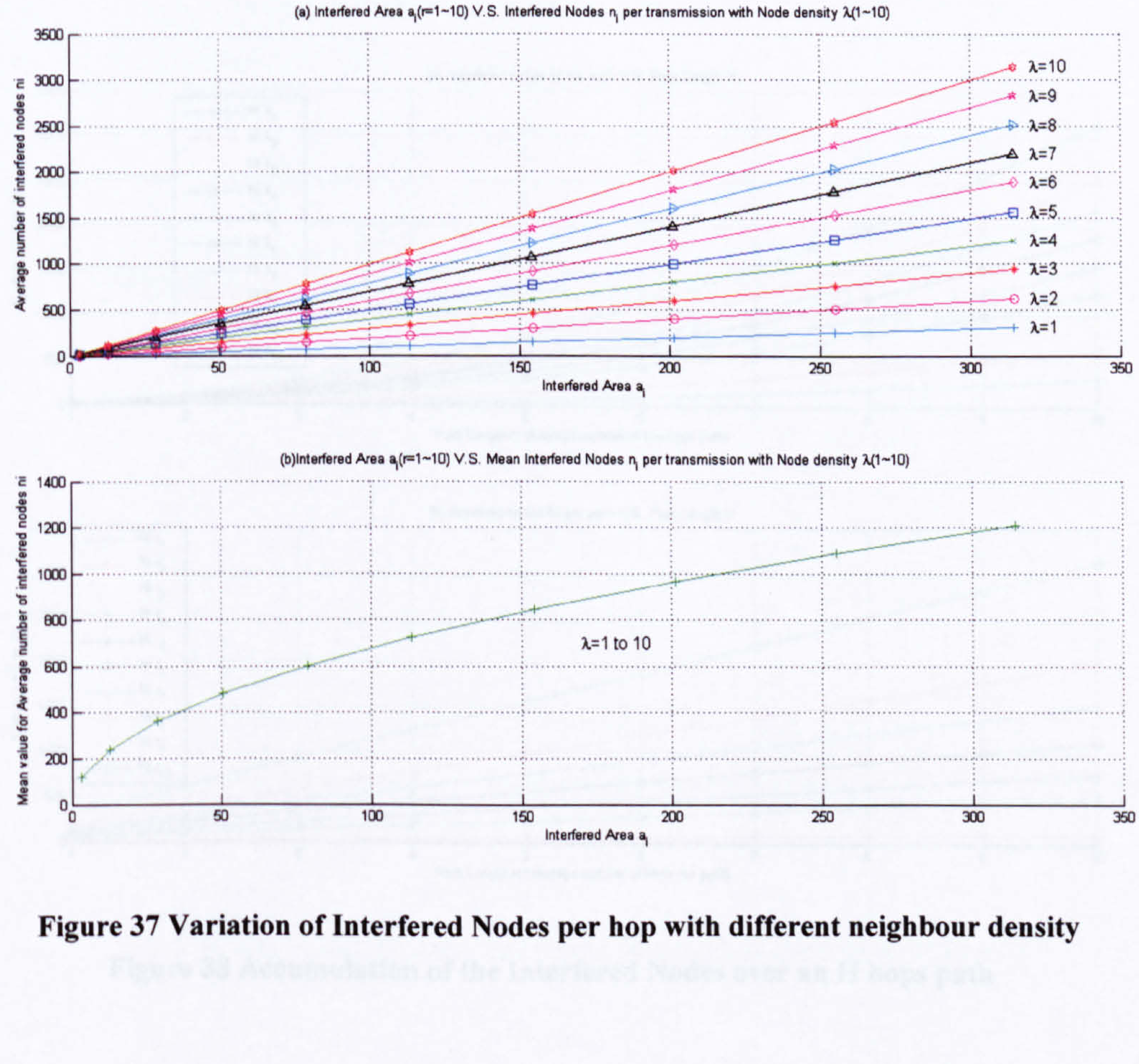
Figure 36 (a), (b) shows the variation and increase on the number of interfered node per hop, $\overline{n_i}$, according to different link length, *r*, for every constant value of node density, λ.



(a)Interfered nodes per hop vs. Node density λ (1~10) when r=1~10

(b)Interfered nodes per hop vs. Node density λ (1~10) when r=1~10

Figure 36 Variation of Interfered Nodes *ni* per hop with different link length *r*



(a) Interfered Area a_i(r=1~10) V.S. Interfered Nodes n_i per transmission with Node density λ(1~10)

(b)Interfered Area a_i(r=1~10) V.S. Mean Interfered Nodes n_i per transmission with Node density λ(1~10)

Figure 37 Variation of Interfered Nodes per hop with different neighbour density

Figure 37 (a), (b) shows the variation on the *number of interfered* node per hop $\overline{n_i}$ according to different node density λ for each constant value of link length r. In terms of ***delay***, assuming the transmission, queueing, processing, and propagation time of all the transmissions are the same, then the same amount of traffic sent through a path will experience longer delays in a multihop transmission system, and the average end-to-end delay $\bar{x}$ will be H times of the single-hop delay $\bar{x}_i$.

$$\bar{x} = \bar{x}_i \times H \tag{5.7}$$

In this sense the longer the path length, the longer the delay in multihop transmission style, and the lower per node throughput or channel utilization for processing the local node's own traffic, due to the relaying burdens and avalanche effects.

If a large number of interfered nodes is accumulated over an H hop path, this means the relaying burden along the path is also high. On the other hand, the throughput is low. The accumulation of N_i over a number of hops H is shown in Figure 38, where (a) shows the N_i growth in number over H hops with a different neighbour node density λ, and (b) is with different link length r. This matches the phenomenon described earlier with equation 5.5.

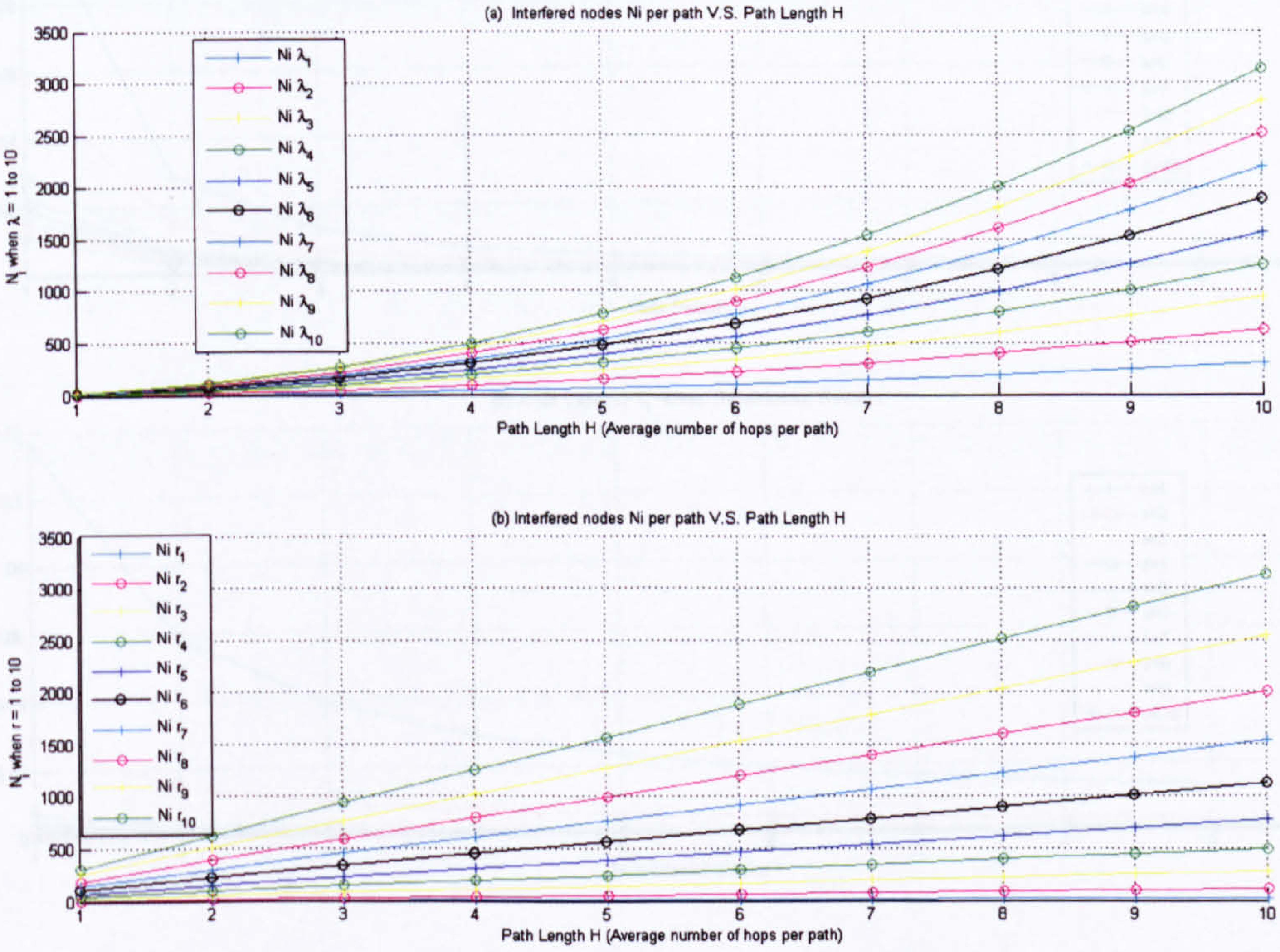


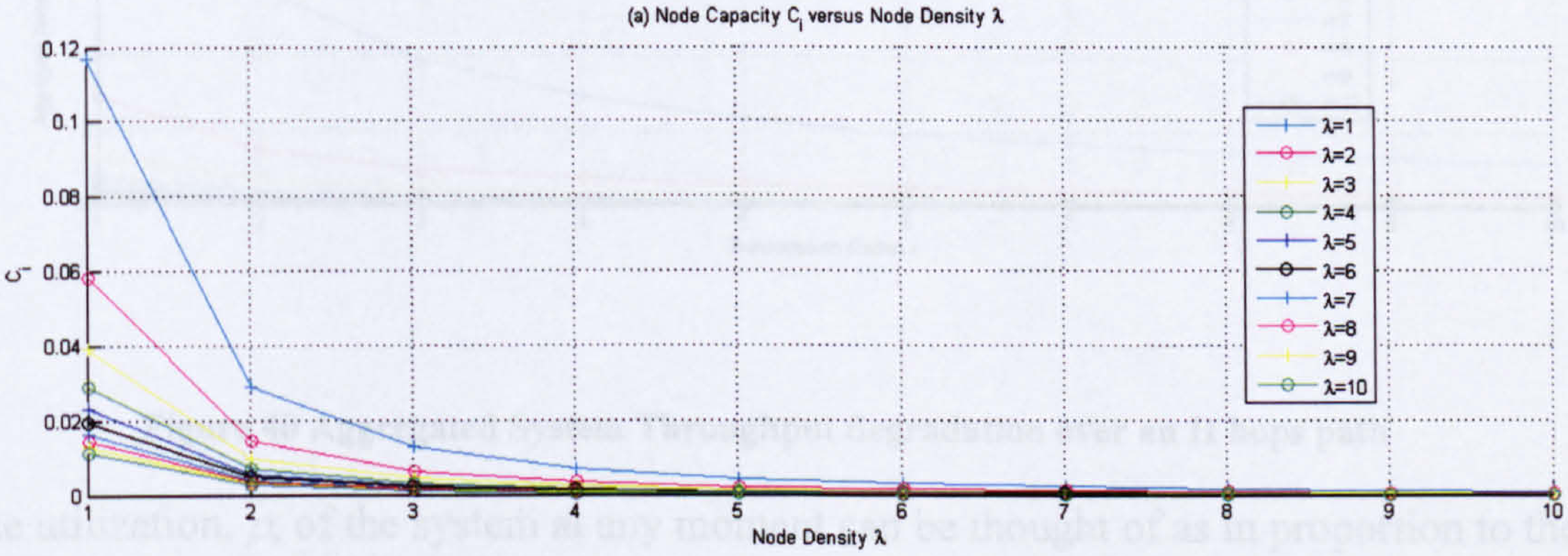
Figure 38 Accumulation of the Interfered Nodes over an H hops path

Substituting equation 5.7 for 5.6, we get the equation shown in 5.8 and 5.9 for the average per-node throughput, which is in proportion to the number of hops, H, and neighbouring node density, n_i. These two measures represent the influence of relaying burden and the avalanche effect of flooding on the per-node throughput C_i.

$$C_i = Multihop \quad node \quad Throughputs = (\lambda(s) \cdot \bar{x} \cdot H)\frac{1}{n_i} \tag{5.8}$$

$$C_i = Single-hop \quad node \quad Throughputs = (\lambda(s) \cdot \bar{x})\frac{1}{n_i} \tag{5.9}$$

The per-hop throughput C_i is under the strong influence of the per-hop interfered node n_i, and, if the neighbouring node density is constant, the link length, as shown separately in Figure 39 (a) and (b). In other words, a higher neighbour node density results in lower per-hop throughput; or with the same neighbour node density, a longer link length will also contribute to a larger number of neighbouring nodes. Consequently, the more neighbours sharing the same resource, the lower the amount of throughput a node could deliver.


(a) Node Capacity C_i versus Node Density λ

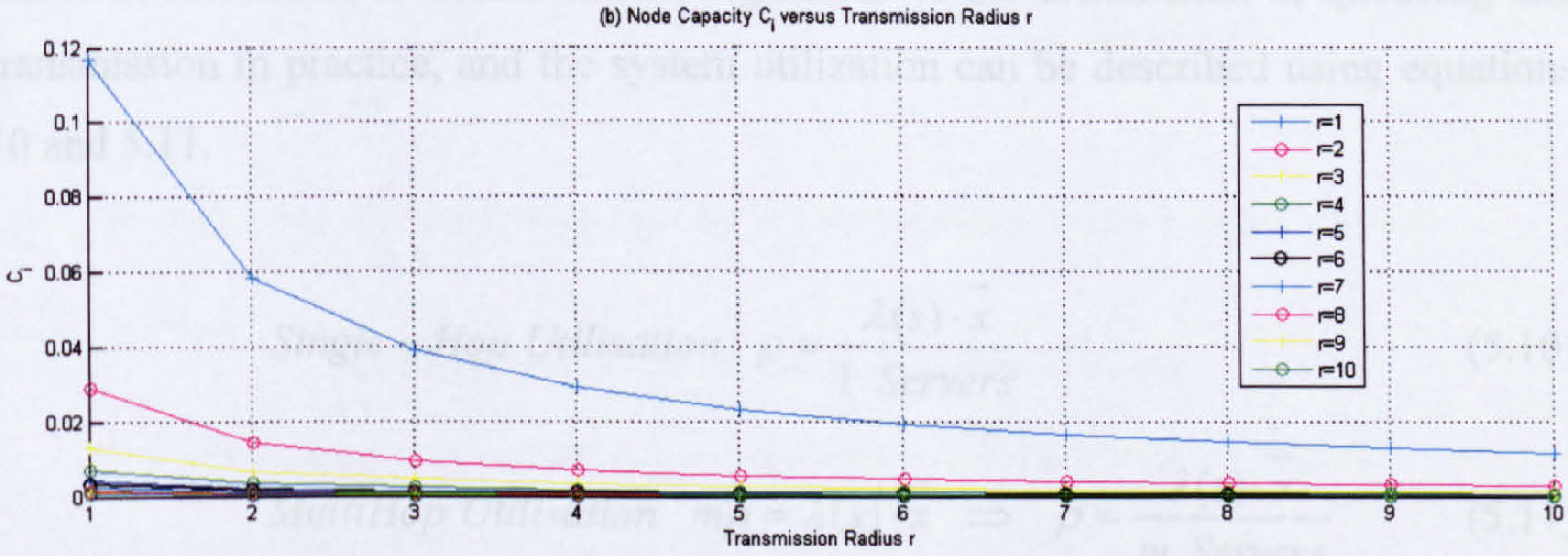
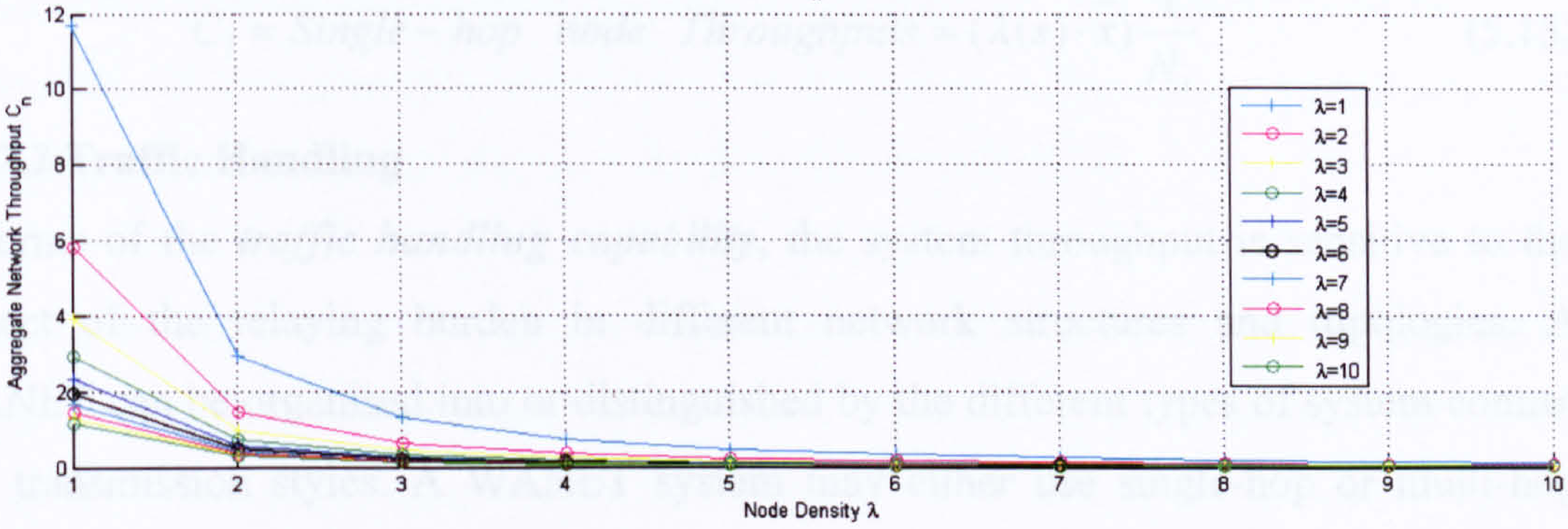

(b) Node Capacity C_i versus Transmission Radius r

Figure 39 Degradation of the Per Node Throughput over an H hops path

The same cause may also affect the system throughput, C_n. If the average per-hop throughput C_i decreases, the aggregated system throughput will decrease accordingly, in a case where the system operates with little or no spatial reuse. The outcome of this effect on C_n is show in Figure 40.

(a) Network Throughput C_n versus Node Density λ



(b) Network Throughput C_n versus Transmission Radius r

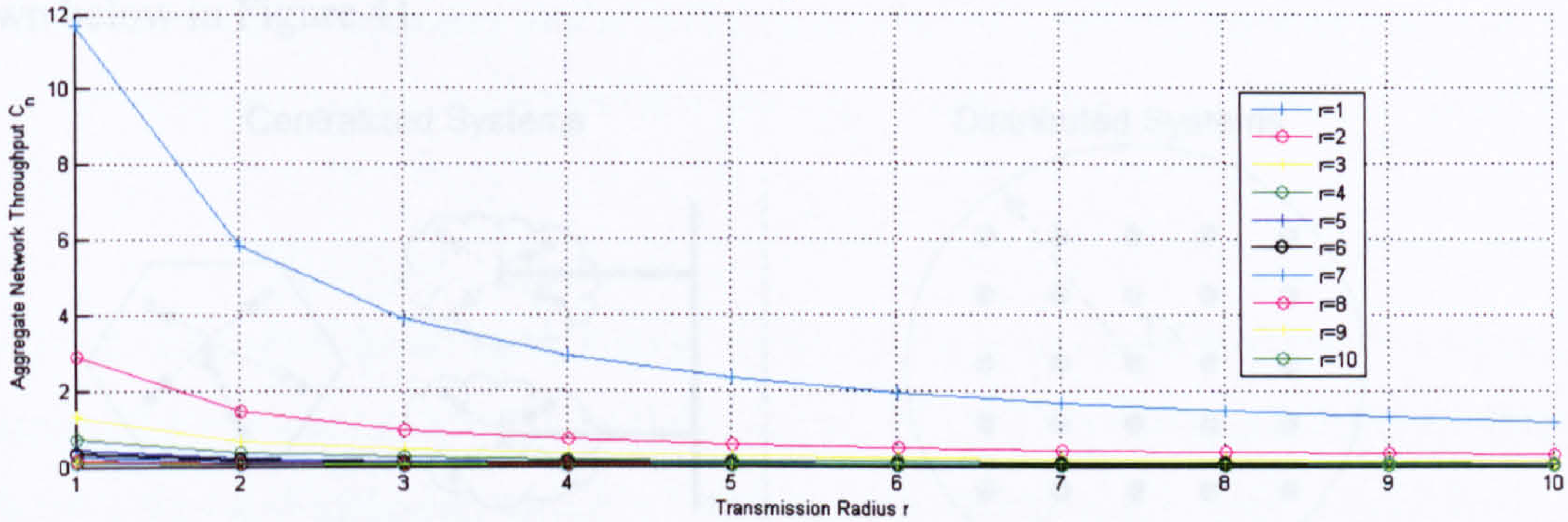


Figure 40 Aggregated System Throughput degradation over an H hops path

The utilization, ρ, of the system at any moment can be thought of as in proportion to the number of simultaneous transmissions, regardless of the distribution of queueing and retransmission in practice, and the system utilization can be described using equations 5.10 and 5.11.

$$Single-Hop\ Utilisation\quad \rho = \frac{\lambda(s)\cdot\bar{x}}{1\ Servers} \tag{5.10}$$

$$MultiHop\ Utilisation\quad m\rho = \lambda(s)\cdot\bar{x}\ \Rightarrow\ \rho = \frac{\lambda(s)\cdot\bar{x}}{m\ Servers} \tag{5.11}$$

The utilization normally represents the throughput or server busy period in the percentage for assessing the system performance, hence equations 5.7 and 5.8 can be rewritten into equation 5.12 and 5.13 with the same assumptions.

$$C_i = Multihop\ \ node\ \ Throughputs = \frac{\lambda(s)\cdot\bar{x}\cdot H}{m}\frac{1}{N_i} \tag{5.12}$$

$$C_i = Single-hop\ \ node\ \ Throughputs = (\lambda(s)\cdot\bar{x})\frac{1}{N_i} \tag{5.13}$$

5.2.2.2 Traffic Handling

In terms of the ***traffic handling capability***, the system throughput is sensitive to the impact of the relaying burden in different network structures and topologies. A WANET can be organised into or distinguished by the different types of system control and transmission styles. A WANET system may either use single-hop or multi-hop transmissions, and manages routing in a centralized or distributed control manner, as shown below in Figure 41.

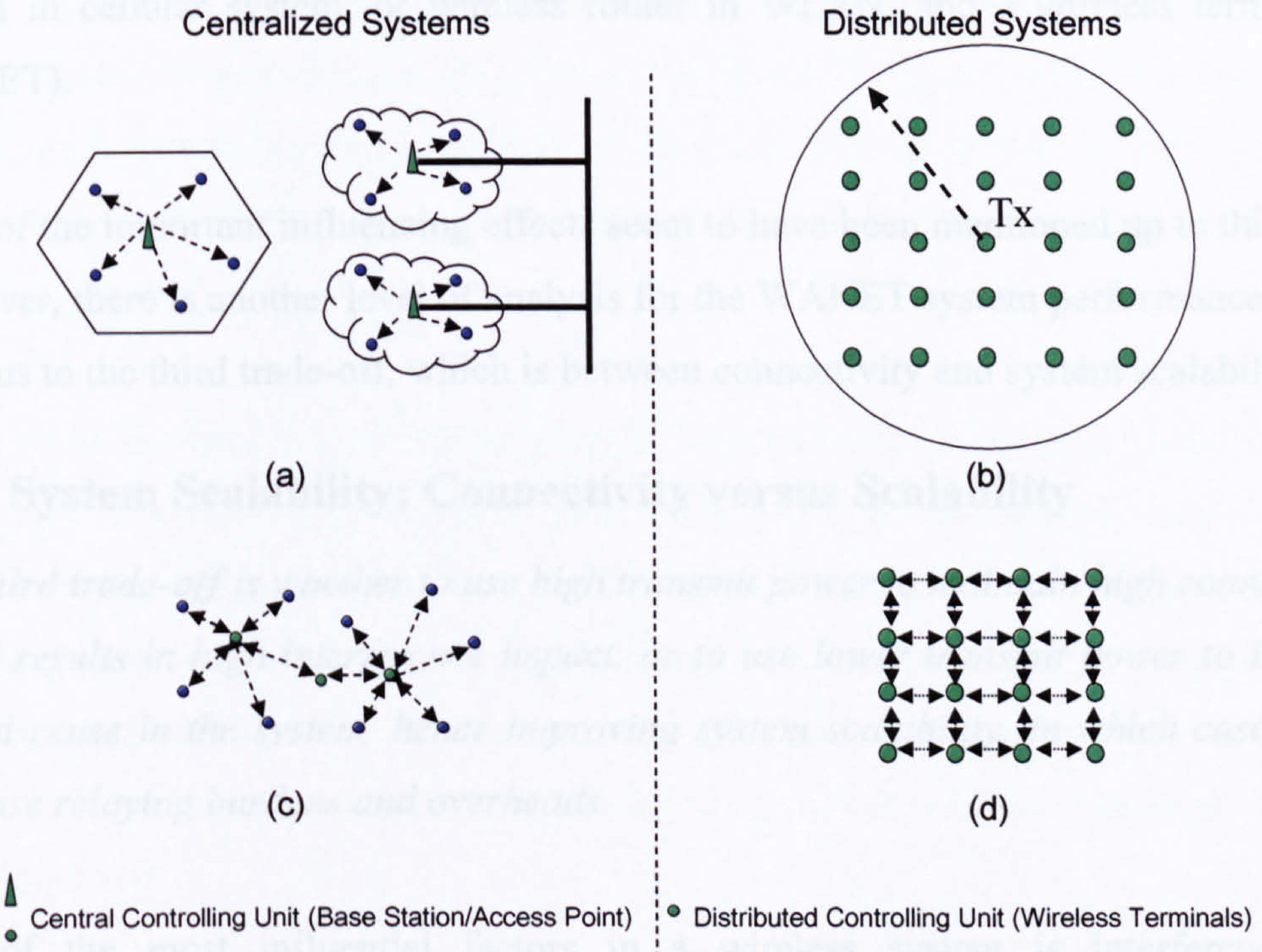


Figure 41 Network architectures in wireless networks

a) Centralized single-hop routing, e.g. conventional or cellular system, or WLAN;

b) Distributed single-hop routing, e.g. high transmit power WANET routing;

c) Centralized multi-hop routing, e.g. proactive or hierarchical WANET routing;

d) Distributed multi-hop routing, e.g. reactive WANET routing.

Under different network structures, *traffic congestion* and *transmission collisions* can seriously damage system throughput and accelerate the accumulation of relaying delay and burdens, consequently paralyzing system traffic. Centralized control, as shown in Figure 41 (a) and (c), in either single-hop or multihop transmission systems, will attract *aggregated traffic* towards a central controlling point, which results in aggravating *transmission contention* and *collision* problems during the higher local multiple access contention. Conventionally, improving the hardware capacity, or optimizing the queueing or scheduling scheme, may resolve traffic bottlenecks.

This means that for any wireless terminal that schedules multiple access demands, a higher surrounding neighbouring node density results in a higher number of access demands at each hop, and consequently increases the *probability of collision.* On the contrary, the distributed control, as shown in Figure 41 (b) and (d), decentralizes traffic and can reduce the *processing and scheduling burden* of a controlling unit (i.e. a base-station in cellular system, or wireless router in WLAN, and a wireless terminal in WANET).

Most of the important influencing effects seem to have been mentioned up to this point. However, there is another level of analysis for the WANET system performance, which leads us to the third trade-off, which is between connectivity and system scalability.

5.2.3 System Scalability: Connectivity versus Scalability

*The **third trade-off** is whether to use high transmit power to maintain high connectivity, which results in high interference impact, or to use lower transmit power to increase spatial reuse in the system, hence improving system scalability, in which case it may increase relaying burdens and overheads.*

One of the most influential factors in a wireless system is interference and, consequently, its impact on system performance. In telecommunication systems, scalability is a desirable property in a system that indicates the ability to handle growing numbers of tasks as the system size and user population grow [105]. However, problems such as *co-channel interference*, *relaying overheads and delays*, *extra control information*, *contention collisions and congestions* may seriously decrease the scalability of a WANET system.

A transmitting node interfering with the neighbouring co-channel and adjacent-channel users, and the resulting impact on system performance, is what we refer to as "interference impact". Interference impact is easier to understand with an example of "a conference room full of people", where if everyone speaks loudly at the same time, each individual can hear multiple conversations as, noise rather than any sensible words or sentences. A similar phenomenon may occur in a wireless system, when users or terminals transmit their signals via a shared frequency channel. If all terminals transmit at high power (implying loud speaking) and try to make their own message clear to the receiving end, then the communication air interface (equivalent to the conference room) will fill with relaying and control overheads, noise or interference, and messages delivered using the same shared channel may collide at the receiving end.

Conventional cellular systems resolve this problem by introducing *frequency reuse*, so that the overheads, interference, contention, and congestions, can be dealt with locally and separately by a stationary base-station. In WANET, *spatial reuse* combined with localized (distributed) control could improve scalability by physically separating two simultaneously transmitting co-channel terminals, and manage the interfering factors mentioned above.

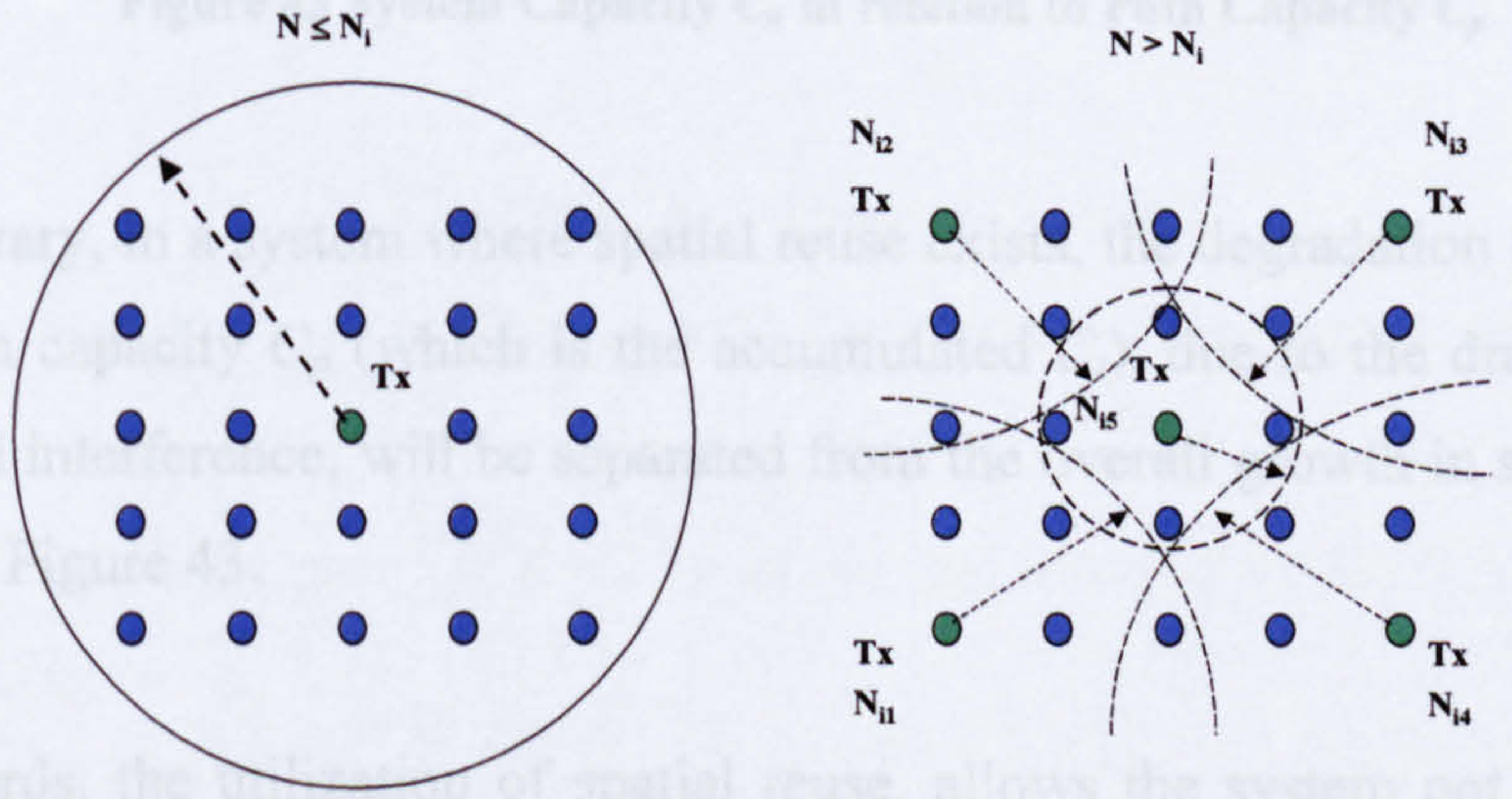
A safe reuse distance is a critical measure in both frequency reuse and spatial reuse. In both cases this safe reuse distance makes sure that two co-channel nodes that are simultaneously transmitting are not interfering with each other. The difference is that frequency reuse in the cellular case has a safe reuse distance that is pre-planned for stationary base-stations; the spatial reuse in WANET can only coordinate between terminals to make sure co-channel nodes do not transmit within a safe reuse distance neighbourhood.

In general, a higher spatial reuse ratio in the WANET system means the scalability for this system is higher. Thus minimizing the transmission link length allows a higher ratio of spatial reuse in a distributed system, however coordinating between terminals is difficult due to the lack of centralized controls.

If the spatial reuse ratio between the wireless terminals is optimally arranged and coordinated, then these previous arguments all have a solution to themselves. By measuring and finding the optimal ratio of spatial reuse, the measures that influence system performance, such as the *optimal number of hops* along a path, *optimal number of relays*, *optimal relaying delays*, *optimal number of transmission demands*, and *minimum interference from neighbouring nodes*, can all be achieved or improved.

$$S = \frac{N}{n_i} \tag{5.14}$$

Assuming that each routing task in the WANET has the same path length, in H hops, then each node that is uniformly located uses the same level of transmit power, the overall system user population, denoted by N, is in proportion to the average number of neighbouring nodes per hop, denoted by n_i. The number of simultaneous transmissions that relate to the spatial reuse ratio, denoted by S, can be measured using following equation. This equation 5.14 holds in time division systems, whereas in frequency and code decision systems may vary, due to the different scheduling mechanisms. Figure 42 (a) shows the neighbouring node n_i in a single-hop system, and (b) shows N_i in a multihop system, where a number of simultaneous transmissions S takes place.



(a) $N_i = N$

(b) $N_i = N_{i1} = N_{i2} = N_{i3} = N_{i4} = N_{i5}$

Figure 42 Neighbouring Node Densities (a) Single-hop (b) Multihop

If we consider that the local per-node capacity or throughput will decrease eventually to zero along a multihop path, as either the neighbouring node density or the access demand grows, then the overall system throughput will definitely decrease to zero if no spatial reuse is introduced to separate or reduce the combined effects of relaying burdens, interference impact and contentions.

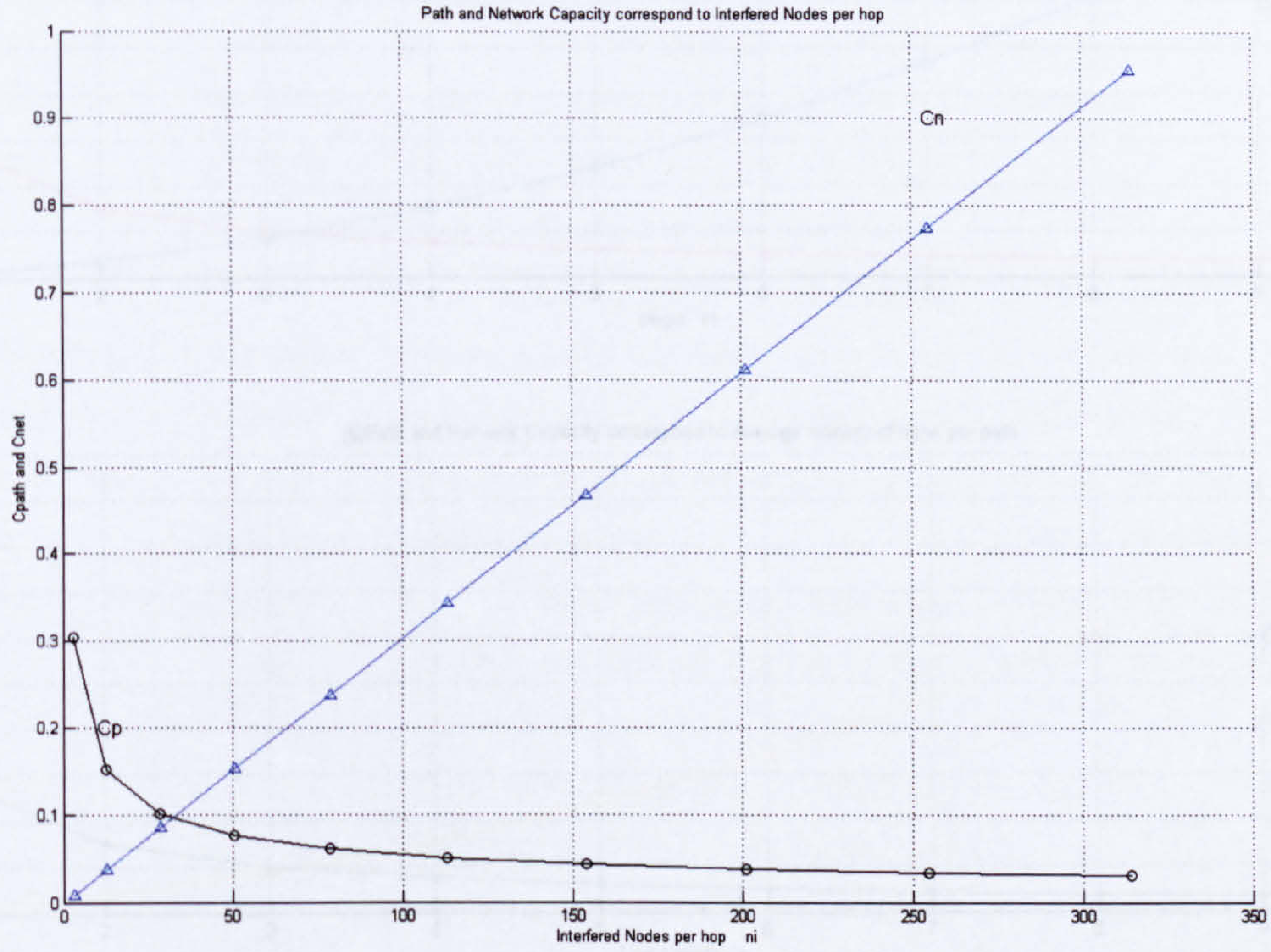


Figure 43 System Capacity C_n in relation to Path Capacity C_p

On the contrary, in a system where spatial reuse exists, the degradation of a per node C_i and per-path capacity C_p (which is the accumulated C_i), due to the dragging effect of relaying and interference, will be separated from the overall growth in system capacity, as shown in Figure 43.

In other words, the utilization of spatial reuse, allows the system not only to be less influenced by the relaying and controlling burdens to multiple points, but also reduces the interference and contention by distributing these impacts to multiple operating servers. The outcome of this theory stands in the case of multihop path, using different link lengths as shown in Figure 44 (a) and (b) respectively.

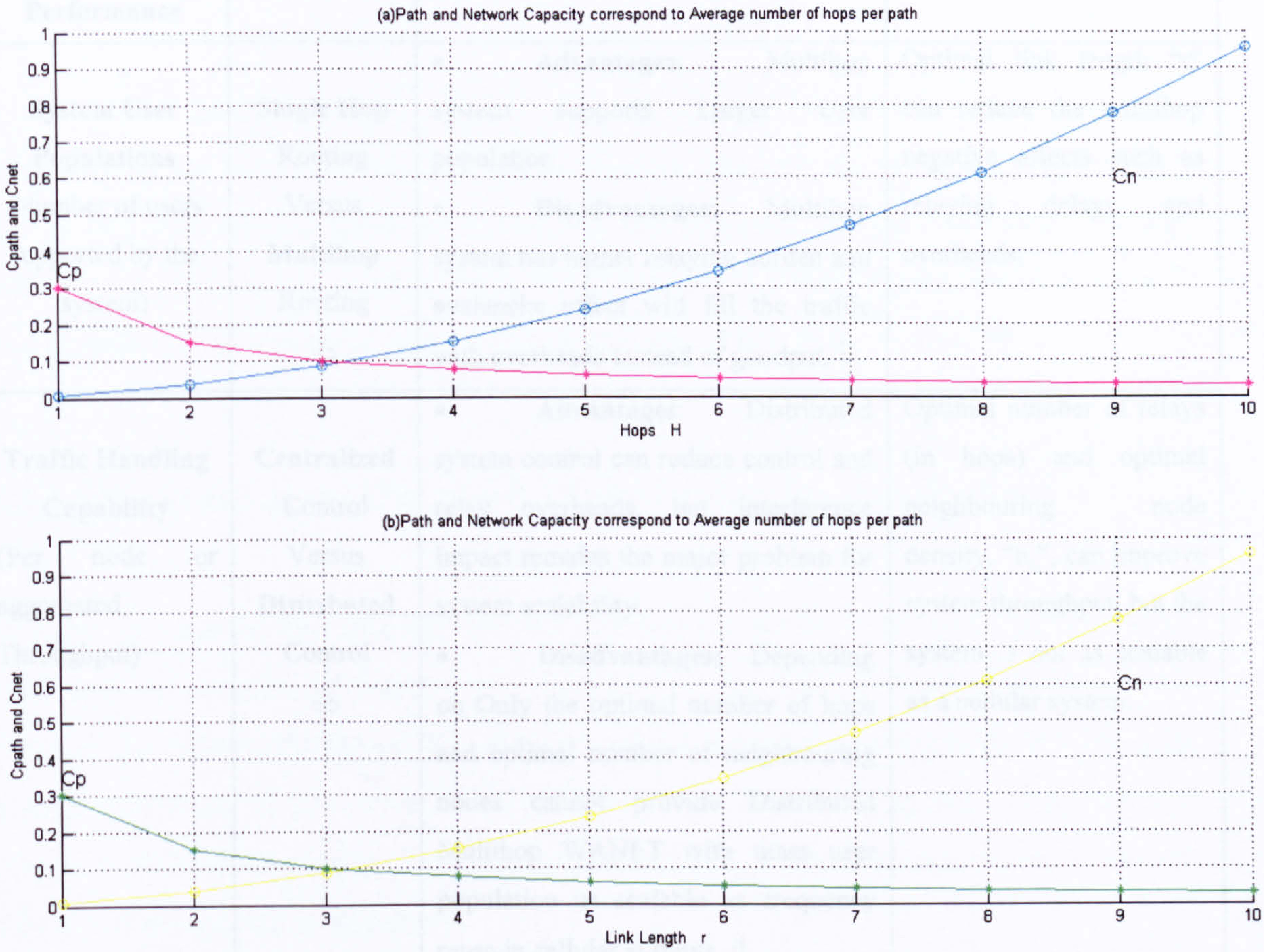


Figure 44 The variation of C_n and C_p

(a) Variation over an H hop path (b) Variation when link length is different

Table 5 presents a summary of the key issues and inter-connections in the analysis of WANET system.

WANET system Performance	**Trade-Offs**	**Conclusions**	**Solutions**
System User Populations (Number of users supported by the system)	**Single Hop** Routing Versus **Multihop** Routing ⇒	• **Advantages:** Multihop system supports Larger User population • **Disadvantages:** Multihop system has higher relaying burden and avalanche effect will fill the traffic with overheads instead of goodput. ⇓	Optimal link length “r” can reduce the multihop negative effects such as relaying delays and overheads.
Traffic Handling Capability (Per node or aggregated Throughput)	**Centralized** Control Versus **Distributed** Control ⇒	• **Advantages:** Distributed system control can reduce control and relay overheads, but interference impact remains the major problem for system scalability. • **Disadvantages:** Depending on Only the optimal number of hops and optimal number of neighbouring nodes cannot provide Distributed Multihop WANET with mass user population as scalable as frequency reuse in cellular systems. ⇓	Optimal number of relays (in hops) and optimal neighbouring node density, “n_c”, can improve system throughput, but the system is not as scalable as a cellular system.
System Scalability (asymptotical system capacity which barely exceed 1 Erlang)	Connectivity Versus Scalability ⇒	• **Advantages:** Multihop Distributed system can achieve higher capacity and reduce control and relay overheads by balancing between Spatial Reuse and Interference Impacts. • **Disadvantages:** Distributed System has highly complicated behaviour, thus difficult to scale (i.e. lower scalability). ⇓	Optimal spatial reuse ratio can improve scalability in some scenarios, and dynamic spectrum sensing can provide a dynamic distributed frequency reuse that can completely resolve the interference impact problem

Table 5 Summaries of the Three Trade-Offs

5.3. Strategies for System Capacity Enhancement

In the above analysis, some of the most influential factors in system capacity enhancement are identified and studied. A *minimum impact routing* (MIR) mechanism that is designed around these measures, e.g. interference impact, local neighbouring node density, and spatial reuse, is being developed. This remaining section in chapter 5 will present the fundamental elements or techniques of this MIR mechanism.

5.3.1. Physical Layer: Distributed Interference Impact Probing

In order to measure and quantify the interference impact in a WANET environment, MIR uses a *Distributed Interference Impact Probing* (DIIP) mechanism can both quantify the *impact of interference* and detect the local *neighbouring node density* with a single measure of *Disturbed Nodes* (DNs). This DNs probing can be a periodic detection of the dynamically changing neighbouring node density, or it can be a one-off measure if the system is stationary.

This mechanism gives each node two pieces of important information: one is the quantified interference impact measure, DNs, which later can be used for making a routing decision as its criteria depend on the impact situation in the network; the other one is the local neighbouring node density reflected also by DNs, achieved by comparing actual measured DN with a pre-defined optimal neighbour node density threshold, maximum DNs (DN_{max}). The transmitting node can adjust its transmit power with respect to DN to this DN_{max} measure to adjust the area of interference impact it has made, as well as control the number of neighbouring nodes that suffer this impact.

5.3.2. Access Control Layer: Distributed BTMA Scheduling

The MIR routing mechanism requires a multiple access control (MAC) protocol model to deal with multiple access contention, avoid collisions, and prevent hidden terminal problems.

It was Kleinrock and Tobagi [32, 106] that recognised and analysed the hidden terminal problems in wireless systems and proposed the Busy Tone Multiple Access (BTMA) solution to deal with this problem. The available frequencies are divided into a data channel and a control channel. While a terminal is receiving data on the former channel, it places a busy-tone on the control channel and signals to other potential senders that if the receiver is busy, they should defer their transmissions. This protects the system from normal contention collisions and hidden terminal problems.

BTMA can also eliminate the *exposed terminal* problems if the sender ignores the carrier sense signal when there is no busy-tone on the control channel. The problem with BTMA is that it requires splitting the channel into two, making receivers more complex. The two bands also need to be separated by a guard band, which wastes radio spectrum. Also, since the propagation characteristics of the radio link are dependent on frequency, a terminal might hear just one of the two signals (busy-tone or data)[32, 106].

5.3.3. Network Layer: Distributed Routing

In terms of routing, a distributed **Local Communication Group** (LCG) is defined by each existing node to achieve localized topology control, which reduces the total amount of routing information that is managed by each terminal, as well as reducing the contention intensity by processing only the demand from its one-hop neighbours. The routing information propagation is achieved by distributed *flooding with control* in conjunction with the distributed *asymmetrical routing.*

5.3.3.1 Distributed LCG

A LCG is a cluster of all immediate neighbouring nodes within the transmission range of a terminal. Every node has its own LCG and declares a **Next Forwarding Node** (NFN) to forward data packets toward a known destination in a hop-by-hop manner. Figure 45 shows an example of LCGs and route selection in different types of network topologies (e.g. chain, square, hexagonal, and random) using different routing schemes, e.g. shortest path and MIR. A detailed MIR protocol description is in chapter 6.

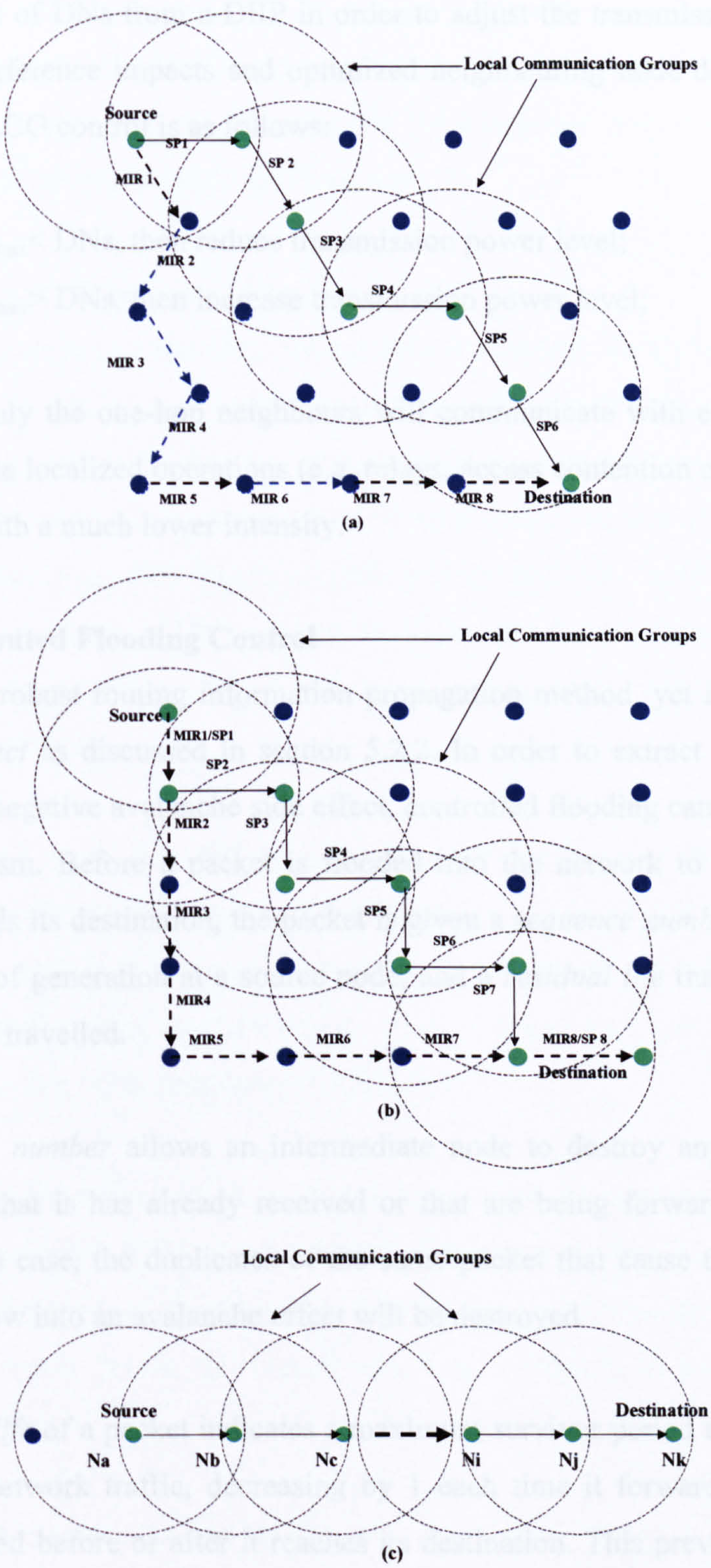


Figure 45 LCGs and Route Selection In 3 different Network Topologies

(a) Hexagon (b) Square (c) Chain

The terminal defines its LCG by referring to the maximum DNs (DN_{max}), which gives a predetermined optimal neighbouring node density. This DN_{max} is then compared to the actual measure of DNs from a DIIP in order to adjust the transmission link length for minimum interference impacts and optimized neighbouring node density. The logical syntax of the LCG control is as follows:

- If $DN_{max} <$ DNs, then reduce transmission power level;
- If $DN_{max} >$ DNs, then increase transmission power level;

In this case only the one-hop neighbours will communicate with each working node, and bring in the localized operations (e.g. relays, access contention control) and routing information with a much lower intensity.

5.3.3.2 Distributed Flooding Control

Flooding is a robust routing information propagation method, yet it has a devastating *avalanche effect* as discussed in section 5.2.2. In order to extract its robustness, and eliminate the negative avalanche side effect, controlled flooding can be used within the MIR mechanism. Before a packet is flooded into the network to be duplicated, and relayed towards its destination, the packet is given a *sequence number*, which indicates the sequence of generation at a source node, and a *residual life* that indicates how far has the packet travelled.

The *sequence number* allows an intermediate node to destroy any duplicates of the same packet that is has already received or that are being forwarded from the same source. In this case, the duplicates of the same packet that cause the relaying loop to eventually grow into an avalanche effect will be destroyed.

The *residual life* of a packet indicates a maximum survivor period that a packet should exist in the network traffic, decreasing by 1 each time it forwarded, and eventually being destroyed before or after it reaches its destination. This prevents a packet being duplicated and relayed forever in the traffic flow. However, within this residual life, a packet could still create multiple duplicates and result in a small scale limited version of the avalanche effect, with a total number of duplicates that are measurable in an idealized network scenario using equation 5.6, $N_t = n_t \cdot n_e^H$, where each transmitting node n_t has an average number of neighbouring nodes n_e at each hop. The number of duplicates, N_i (which also measures the *number of interfered nodes* i.e. the end-to-end accumulated disturbed nodes along a path), is in proportion to the path length H (measured in hops) and the number of neighbouring nodes n_e, and N_i grows exponentially as the number of hops increases.

5.3.3.3 Cross-layer Distributed Power Control

The MIR mechanism requires **Distributed Power Control** (DPC) for each terminal for multiple purposes, or a cross-layer meaning. For the *physical layer*, DPC adjusts the transmission link length, so that any one transmitting node will not disturb or interfere with too many neighbouring nodes on a per-hop basis. For the *MAC layer*, the DPC adjusts its transmission link length according to the comparison of pre-defined $\mathrm{DN}_{\mathrm{max}}$ with the actual measure of DNs per hop, to make sure that the transmitting node has a near optimal neighbouring node density, and consequently, each node only handles local access requests.

For the *network layer*, the DPC is responsible for constraining the routing operation through the use of only the local routing information within a LCG instead of global. In this case, each node will process less routing control overheads and, at the same time, reduce the overall routing overheads in the system, due to each node only managing the routing information of its neighbours within its LCG.

From a system level point of view, distributed power control operates in conjunction with DIIP, distributed BTMA, and distributed routing, creating a mechanism that can maintain the balance of transmission link length, neighbouring node density, and interference impact levels. Chapter 6 and 7 will explain in more detail how the mechanism operates.

5.3.3.4 Distributed Asymmetrical Routing

Using **Variable Transmit Power** (VTP) results in asymmetrical radio links on a **Forward path** (FP - from source towards destination) and a **Reverse Path** (RP - from destination return to source). Asymmetrical routes can appear in a network where the FP contains all the forward radio links starting from an originating source node towards the paired destination node; the RP contains the radio links in the opposite direction. This means the one-hop distance on a forward path may appear to be two-hops on the reverse path, in which case it might confuse routing algorithms that use hop-counts as a routing constraint, as shown in Figure 46.

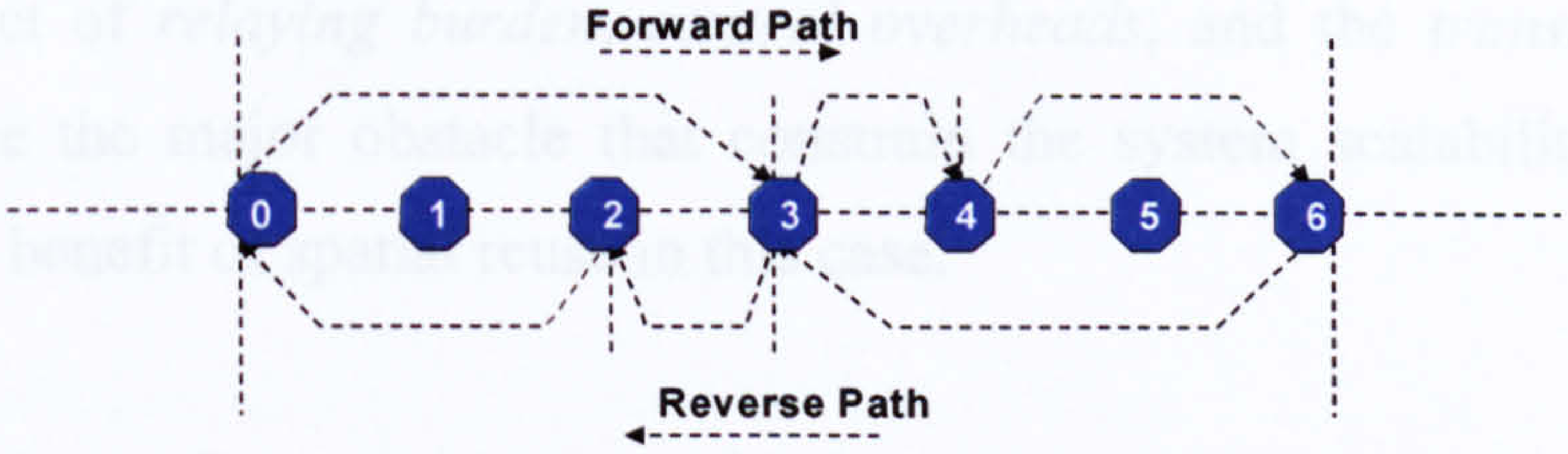


Figure 46 Paths with Asymmetric Links

Numerous strategies have been developed to establish connections on asymmetric routes, as discussed earlier in section 3.4 [96-100]. Route discovery information is disseminated separately for each direction (FP or SP). Previously proposed methods ensure the connectivity relay on bidirectional routing with frequent acknowledgement exchanges.

This results in complex routing operations generating redundant routing overheads. The asymmetrical algorithms could provide a solution to establish the most appropriate routes through a network. WANET routing calls for routing algorithms tailoring the transmit power and routes through the network to fit in with desired constraints (e.g. hop count, battery life, interference) [96-100].

5.4 Conclusion

The early section (5.2) of this chapter identified the key influential factors in enhancing WANET system capacity and scalability, analysing the interrelationship between these factors through the discussion of the three trade-offs.

The first trade-off of whether using single-hop or multihop transmission for routing brings out the desirable feature of multihop transmission, enables simultaneous transmissions to take place, but has drawbacks such as relaying burden. The second trade-off continues the discussion of relaying burden from a centralized versus distributed control aspect, to illustrate the desirable feature of distributed control, in terms of distributed relaying, controlling, and processing burden, among a smaller number of neighbouring nodes. The third trade-off continues with the discussion of localized information processing, further explaining the interference impact. This is the combined effect of *relaying burden, control overheads*, and the *transmission block-outs*, which are the major obstacle that constrain the system scalability, in addition, identifying the benefit of spatial reuse in this case.

In the later section (5.3), a combination of different techniques, which are each aimed at solving one aspect of the system scalability problem, are proposed and these contribute to the construction of a new routing mechanism, the Minimum Impact Routing (MIR) algorithm, which will be discussed in more detail in the next two chapters.

The MIR mechanism senses the interference impact existing in a network, and then constructs and maintains a localized communication group (LCG), where routing information is handled and forwarded with careful *controlled flooding*, to limit the avalanche effect. All these operations are more or less achieved using variable transmit power, which results in a side effect, known as asymmetrical paths, and hence the MIR propagates information using asymmetrical routing.

Spatial reuse in WANET will contribute to the reduction of interference impacts and improve the system scalability. The optimal solution for WANET system scalability is distributed autonomous frequency or spatial reuse, and in this case, it requires dynamically sensing the neighbouring node's spectrum using, and selects non-conflicting channels for communication, in which case it could potentially provide scalability as high as cellular networks, subject only to the impact of relaying burden.

Chapter 6 The Minimum Impact Routing Scheme

Contents

6.1 Introduction

The majority of existing routing algorithms neglect the effect of interference impact and transmission spatial concurrency constraint on the reduction in capacity[17, 33, 36, 37, 70, 107]. This reduction is caused by nodes close to a receiver that are required to be idle to avoid collisions [1, 10, 13, 20]. *The effect of a transmitting node interfering with surrounding nodes while communicating, and the resulting effects on system and user capacity, is what is referred to as "impact".* The throughput furnished to each user will eventually diminish to zero as the number of users is increased [20]. There is a limit on the number of users that can simultaneously communicate in the network, due to the interference between the transmitting nodes and the surrounding nodes that share the same spectrum.

While one node is transmitting, nodes sharing the same spectrum close to a receiver are required to be idle to avoid collisions. Therefore neighbouring nodes of the receiver cannot communicate simultaneously. These nodes are considered to be the Disturbed Nodes (DNs) that are affected by interference, and this "lock out" results in decreased system and user capacity. We consider when a source-destination pair communicates any intermediate nodes which receive significant unwanted interference, is a DN, i.e. except for the intended transmit and receive nodes.

Most early proposed WANET routing algorithms focus on enhancing network connectivity, and they emphasize the benefit of power control on energy consumption or simple link length control, to achieve spatial reuse etc. Based on these theoretical algorithms, many simulations assume that links are symmetrical between neighbouring nodes, yet these assumptions neglect the fact that variable transmission power control yields asymmetrical links, and high connectivity using long links results in low spatial reuse. On the contrary, if each node simply uses the minimum link length, then a high ratio of relay overheads will dominate the network traffic, and therefore it is important to investigate how the transmission interference impact reduces spatial concurrency [10].

The Minimum Impact Routing (MIR) algorithm introduced in this chapter, operates based on the minimisation of the accumulated interference impact on multi-hop communications in WANET. It is very important to restrict such impact in WANET, since a transmitting node will block out adjacent nodes that share the same spectrum. MIR is designed to provide a low number of disturbed nodes, and low interference impact, and hence low routing overhead, high routing efficiency, and high system and user capacity. The MIR scheme is then simulated on a test bed and constructed using a network simulator OPNET.

The simulation model constructed and used in this chapter has three basic network topology scenarios: square, hexagonal, and chain. In each of these scenarios, each transmitting node has been given a different number of immediate neighbours, e.g. 4 in square network, 6 in hexagonal network, and 2 in chain network. The simulation model has 32 identical wireless nodes, each capable of transmitting at an identical transmitting power, which can be manually configured with different power levels. The performance of a WANET simulation model using the MIR scheme is then evaluated and the results collected at the end of each simulation, and validated by processing the resulting data into different formats of plot.

6.2 MIR Theoretical Characteristics

This section introduces MIR in three parts: first is the preliminaries of MIR, which defines link length, interference range, disturbed nodes, route weight, beacon packets, and local communication groups; second is the Distributed Interference Impact Probing (DIIP) principle; third is the MIR protocol, which includes routing information dissemination, route length measuring, and the two stages of the routing table related operation, the route discovery stage and the route maintenance stage.

6.2.1 MIR Terminologies

A **Link Length** (LL) is the distance between two adjacent radio nodes as shown in Figure 47. The LL is defined as the farthest edge that a node's minimum transmit power can reach its immediate neighbours, with the packets still being correctly received, which is adjustable according to the needs of transmitting node.

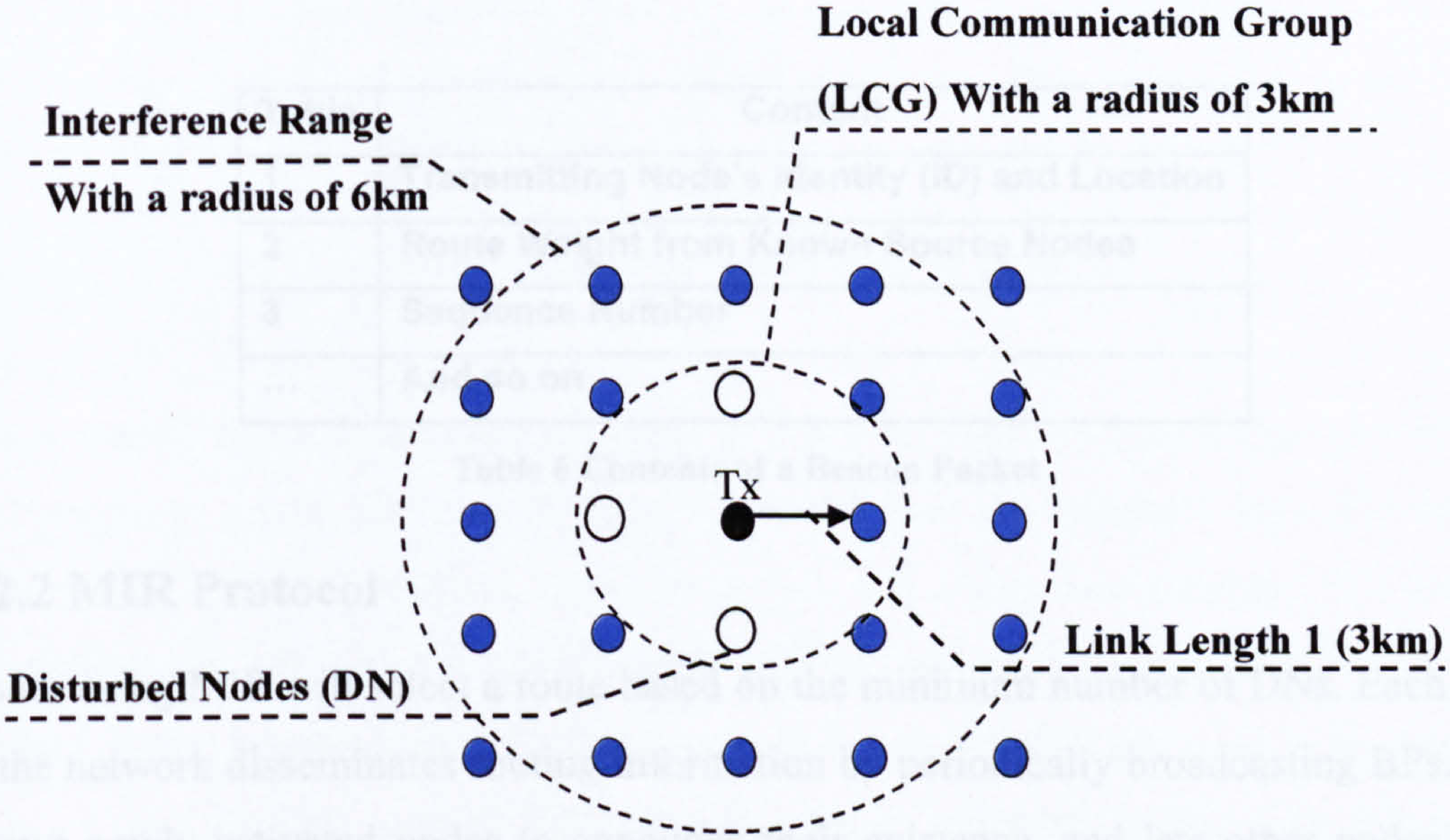


Figure 47 Interference Impact model

Interference Range (IR) indicates the radius of an area where a transmitting node causes an interference impact, which prevents other nodes within the IR to transmit at the same time. The size of the IR is proportional to the LL. Figure 47 shows an example of LL and IR.

Disturbed Nodes (DNs) are all nodes, within an IR, that are receiving an adverse impact. (The DNs are similar to the exposed nodes in WLAN in the sense that they both suffer from transmission block outs, however, in this research, DNs are utilized for not only describing the interfered node, but also measuring interference impact, which we discussed previously in chapter 5)

Route Weight (RW) is the sum of all the DNs along the path between a pair of source and destination nodes.

Beacon Packet (BP) contains sufficient routing information to establish a connection or update other node's routing tables. The BP contains routing information, such as the transmitting nodes' ID, and location. It is broadcast periodically to nodes within the maximum LL range. Nodes receiving a BP update their routing information concurrently, and calculate their distance from the transmitting node. The example content of a BP is shown in Table 6.

Table	Content
1	Transmitting Node's Identity (ID) and Location
2	Route Weight from Known Source Nodes
3	Sequence Number
...	And so on

Table 6 Contents of a Beacon Packet

6.2.2 MIR Protocol

Nodes using MIR will select a route based on the minimum number of DNs. Each node in the network disseminates routing information by periodically broadcasting BPs. This allows newly activated nodes to announce their existence, and lets other nodes learn their Identity (ID) and location etc. By receiving BPs, each node can calculate the distance from itself to the transmitting node and determine its immediate neighbours.

In this way each node will construct a Local Communication Group (LCG), which contains all nodes within a one hop distance, for later transmission of any packets. To communicate with destinations that are outside of one's LCG, a chosen LCG member will forward the packets approximately towards the desired destination by referring to the routing table.

Initially when the routing tables are empty, every node starts to broadcast BPs embedded with a lifetime, so that other nodes can learn of their existence, determine their position, and create entries in each nodes' routing tables. For example, the highlighted column in Table 7 shown below is the entry of N_1 in N_i's routing table (i does not equal 1). After a while every node will have a routing table consisting of one routing entry for each learned source node.

Identified Source Node ID (Source/ Destination)	Forwarding Node ID (Last/next)	Number of Disturbed Nodes (DN)	Sequence Number	Other Routing Control factors
N_1	N_{i-1}	20	015	...
N_2	N_{i-1}	15	203	...
...	...	...	...	...
N_{i-1}	N_i	3	251	...
N_i	N/A	N/A	N/A	...

Table 7 Example of routing table for Node Ni

When a node needs to transmit a packet to a known source node, without knowing a predetermined route, the packet will be flooded into the network, in a hop-by-hop manner with a limited lifetime, (i.e. hop-count), and a sequence number. This ensures the packet will not cause endless retransmissions and unlimited duplication of stale packets. Every intermediate node that forwards this packet will add its own ID, together with the total number of local DNs, into this packet's header. By the time this packet reaches its destination node, its recorded route weight, which is the total number of DNs along the path, will be compared with other previously recorded packets from the same source. The packet with a lowest route weight will be used to update the destination node's routing table, and its last forwarding node's ID will be recorded.

When a node needs to send a packet in a reverse direction, the recorded 'source node ID' will be used as a new 'destination node ID', and the previously recorded 'last-forwarding-node ID' will become the next 'next-forwarding-node ID'. In this way every node will soon determine a Next Forwarding Node (NFN) lead toward each identified node.

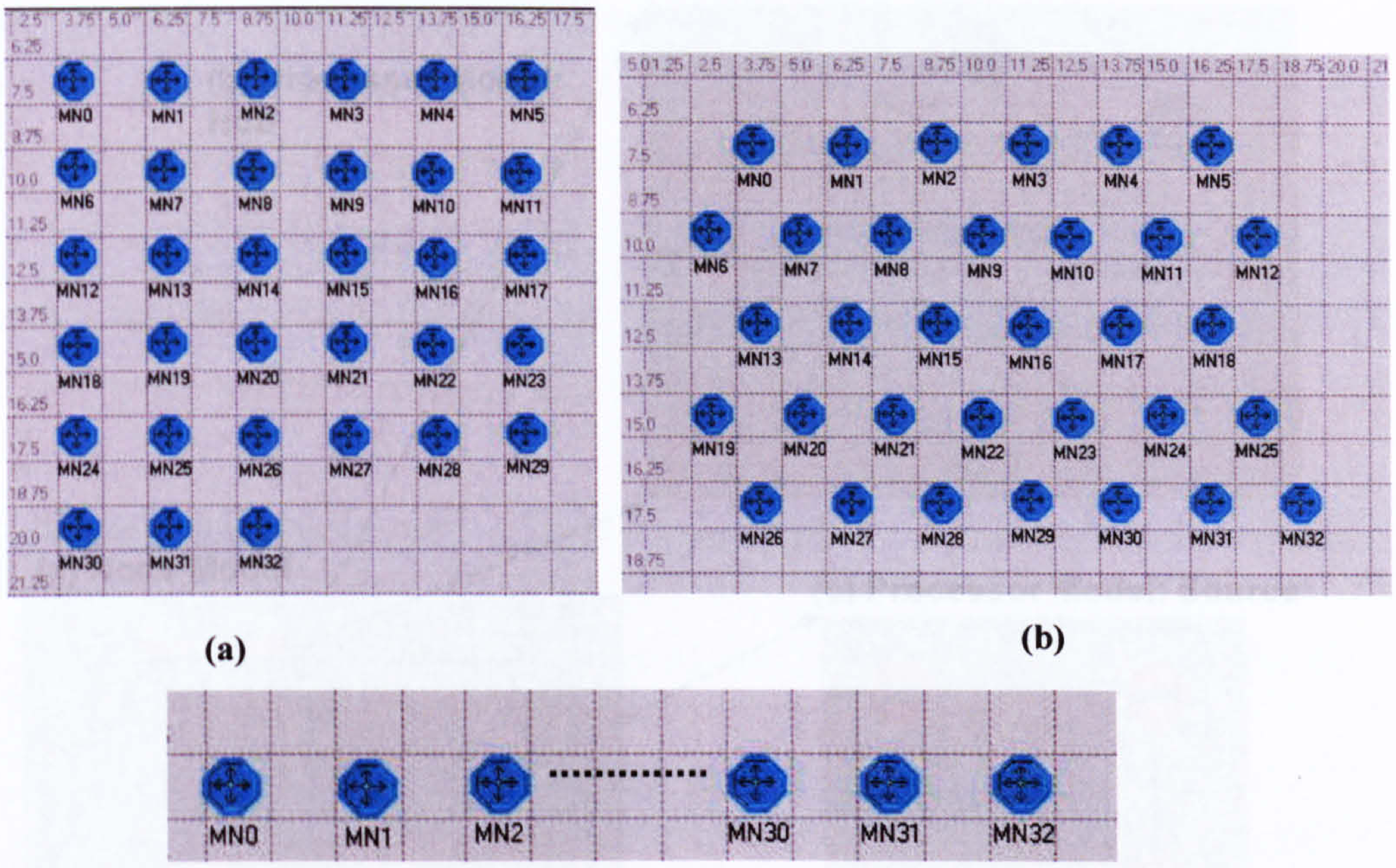
Route maintenance is achieved by updating the forwarding node's ID, using information carried by received packets. Each node forwards a packet to its one-hop away NFN within its LCG. Packets with a valid lifetime are relayed in either the situation where the current node's ID matches the NFN's ID indicated in the packet's header field, or if no NFN ID is specified, then this packet should be broadcast onward. Otherwise, if a received packet is stale or the packet is not addressed to the receiving node, this packet will be discarded. This stops packets exponentially flooding the network and reduces the number of duplicated packets. Packets will time out after their lifetime limit has run out. If changes take place in the network, a node needs to update the routing table with an alternative 'next forwarding node ID' towards a destination node.

6.3 MIR Performance Evaluation

The performance of the WANET model using the MIR algorithm is simulated using the OPNET network simulator. A combination of network scenarios (e.g. hexagon, square, and chain) and link length configurations (e.g. 3,6,9,12 kms) is examined, and then the simulation results are processed using MATLAB.

6.3.1 Simulation Models with Identical node transmit power

The performance of the MIR simulation model was measured, mainly based on the reduction of the total/mean number of DNs, in a unified WANET simulation model with 32 nodes, which organized into three different network topology scenarios, as shown in Figure 48.



(a)

(b)

(c)

Figure 48 Network Topologies

(a) Square Network (b) Hexagonal Network (c) Chain Network

At this stage of the research work, the one-hop distance for all nodes in the network is regulated in different shaped network topologies, in order to connect a different number of neighbours. For instance, within one-hop distance, each node will have a maximum of 6 neighbours in hexagonal topology, 4 neighbours in square topology, and 2 neighbours in chain topology respectively, as shown in Figure 48. These different topologies have been chosen in order to examine the effect of node connectivity on performance.

In all three networks, the internal structure of each wireless node is the same, as demonstrated in Figure 49, where (a) shows the four function blocks (radio receiver-rr, radio-transmitter, hub, and source-src) of a radio node; (b) shows the State-Transition-Diagram (STD) of a processor 'hub', which is the control centre of the node; and (c) is the STD of a source.

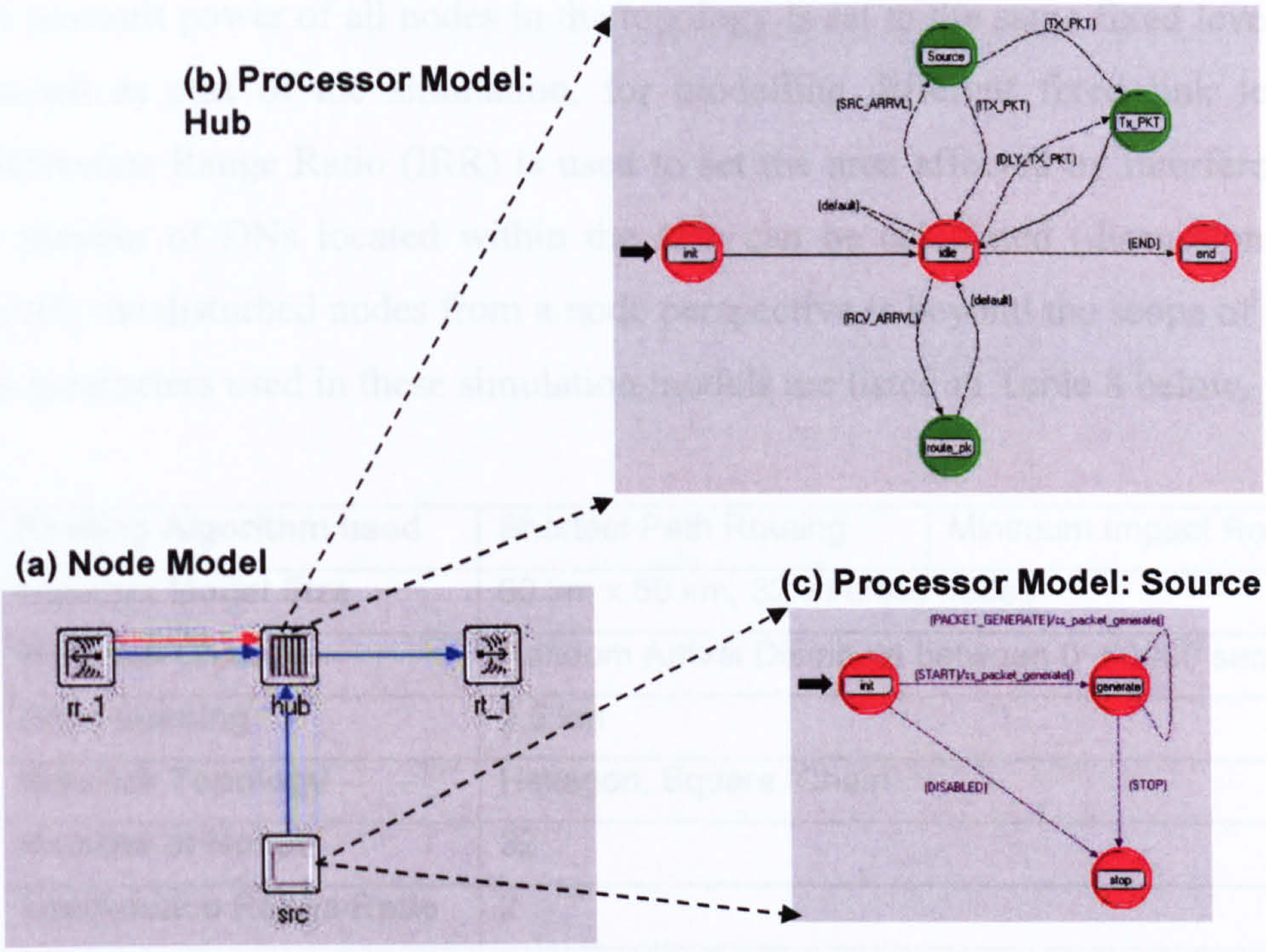


Figure 49 Internal structure example of a radio node

(a) Node Model (b) STD of processor 'hub' (c) STD of Process 'source'

The same network model can be configured to use either the common shortest path algorithm, for simplicity, or be configured to operate using the MIR algorithm, and it is in the STD where the routing algorithms are defined and programmed. Figure 50 shows an abstract logical function structure diagram.

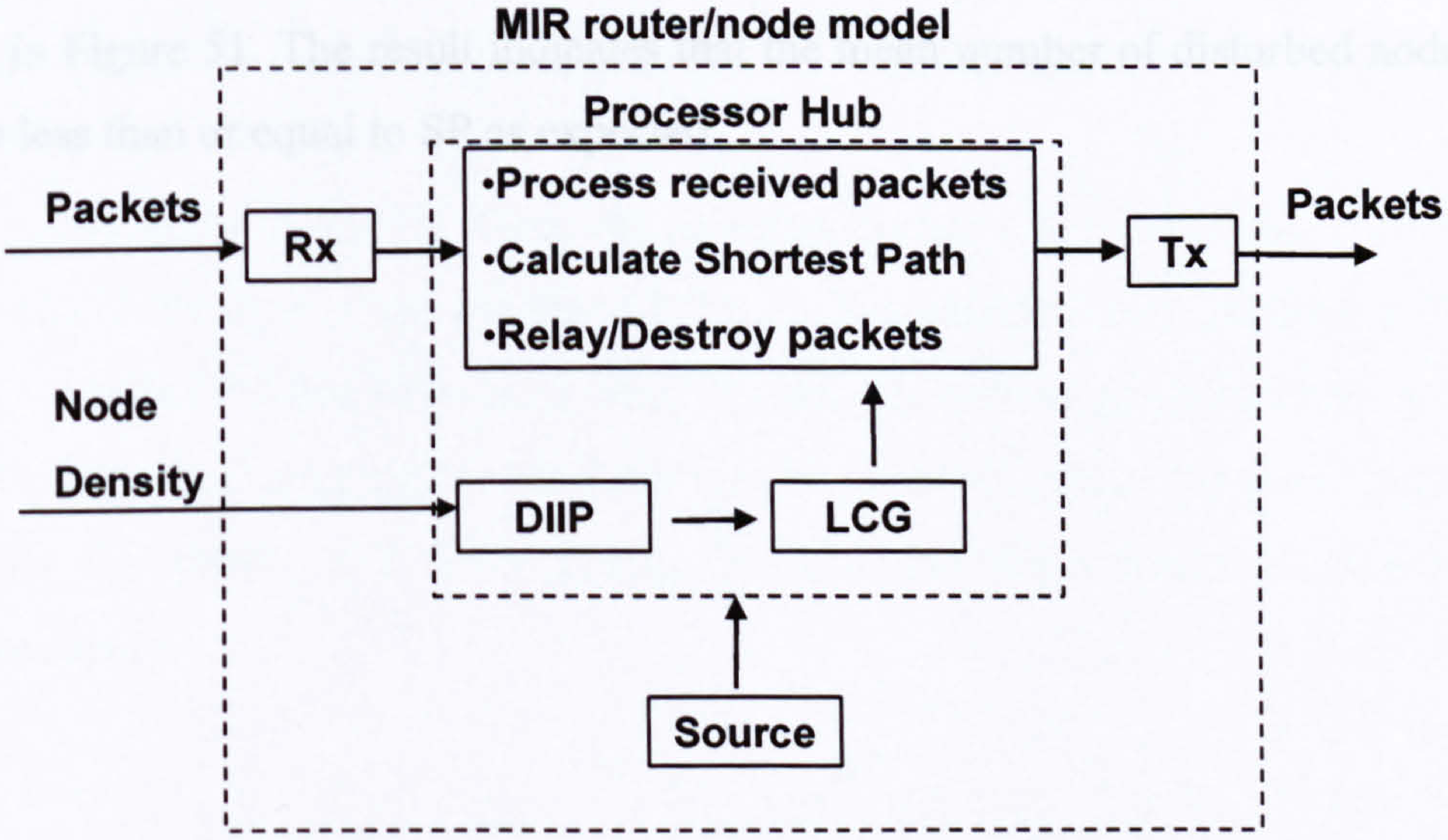


Figure 50 Logical Structure of a MIR router/node

The transmit power of all nodes in the topology is set to the same fixed levels, but later adjusted as part of the simulation, for modelling different fixed link lengths. The Interference Range Ratio (IRR) is used to set the area affected by interference so that the number of DNs located within the area can be calculated (discussion of how to identify the disturbed nodes from a node perspective is beyond the scope of this thesis). The parameters used in these simulation models are listed in Table 8 below.

Routing Algorithm used	Shortest Path Routing	Minimum Impact Routing
Network Model Size	60 km x 60 km; 32 wireless nodes	
Wireless Channel	Random Arrival Distribute between 0~50000 seconds	
Node spacing	2.5 km	
Network Topology	Hexagon, Square, Chain	
Number of Nodes	32	
Interference Range Ratio	2	
Radio Link Length	3,6,9,12,15,18,21 km	
Route Selection Criteria	Shortest Distance (Hops)	Minimum number of DNs

Table 8 Simulation Configurations

6.3.2 Result Evaluation

The simulation results from the network operated under the configuration using Shortest Path (SP) and MIR routing algorithm are compared. Firstly, the mean number of DNs is determined for a hexagonal node topology. The comparison of total number of DNs from SP and MIR with three basic fixed Link Lengths in a hexagon network is shown in Figure 51. The result indicates that the mean number of disturbed nodes with MIR is less than or equal to SP as expected.

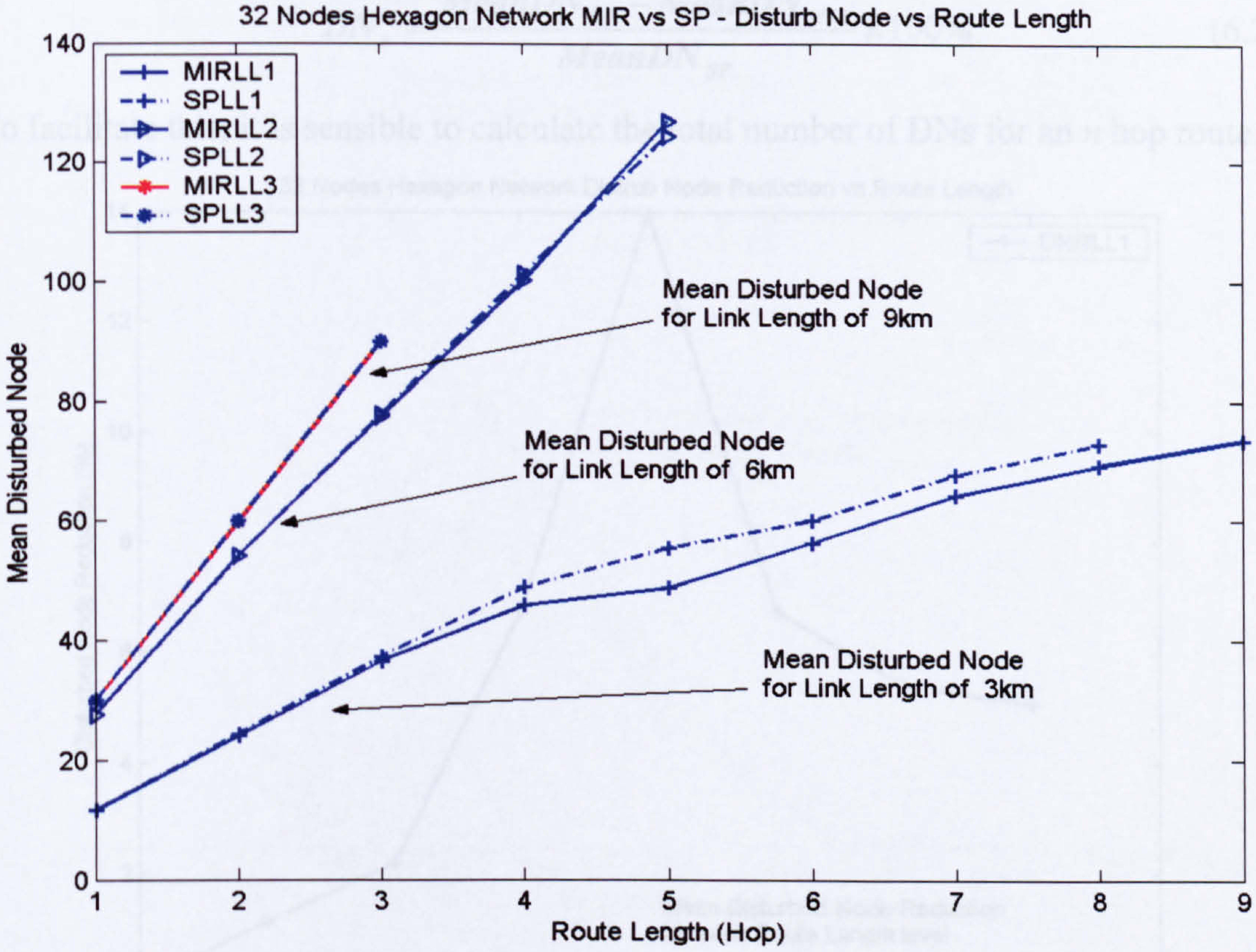


Figure 51 Mean Disturbed Nodes Performance

32 Nodes Hexagon Network SP and MIR Comparison

The difference between SP and MIR is shown more obviously at Link Length level 1 (indexed as SPLL1, MIRLL1), where the one-hop distance is preset to 3 km, and more geographically different routes are possible (in terms of the nodes that get disturbed). This proves in theory that the total number of DNs from MIR is lower than SP, since MIR has chosen the route with lowest number of disturbed nodes.

Link Length levels 2 and 3, where the one-hop distance is 6 and 9 km respectively, shows less difference in the number of DNs. This indicates that the shorter the link length, the better the flexibility with which a node can choose an alternative route using MIR. In this case, reducing the total number of disturbed nodes for each routing task improves the system and user capacity by allowing more nodes to communicate simultaneously.

$$DN_{total} = \sum_{i=1}^{n} DN_i \tag{6.1}$$

It is possible to calculate the mean DN Reduction (DN_r), which is defined by equation 6.2 for investigating the performance of each network topology.

$$DN_r = \frac{MeanDN_{SP} - MeanDN_{MIR}}{MeanDN_{SP}} \times 100\% \qquad (6.2)$$

To facilitate this, it is sensible to calculate the total number of DNs for an *n* hop route.

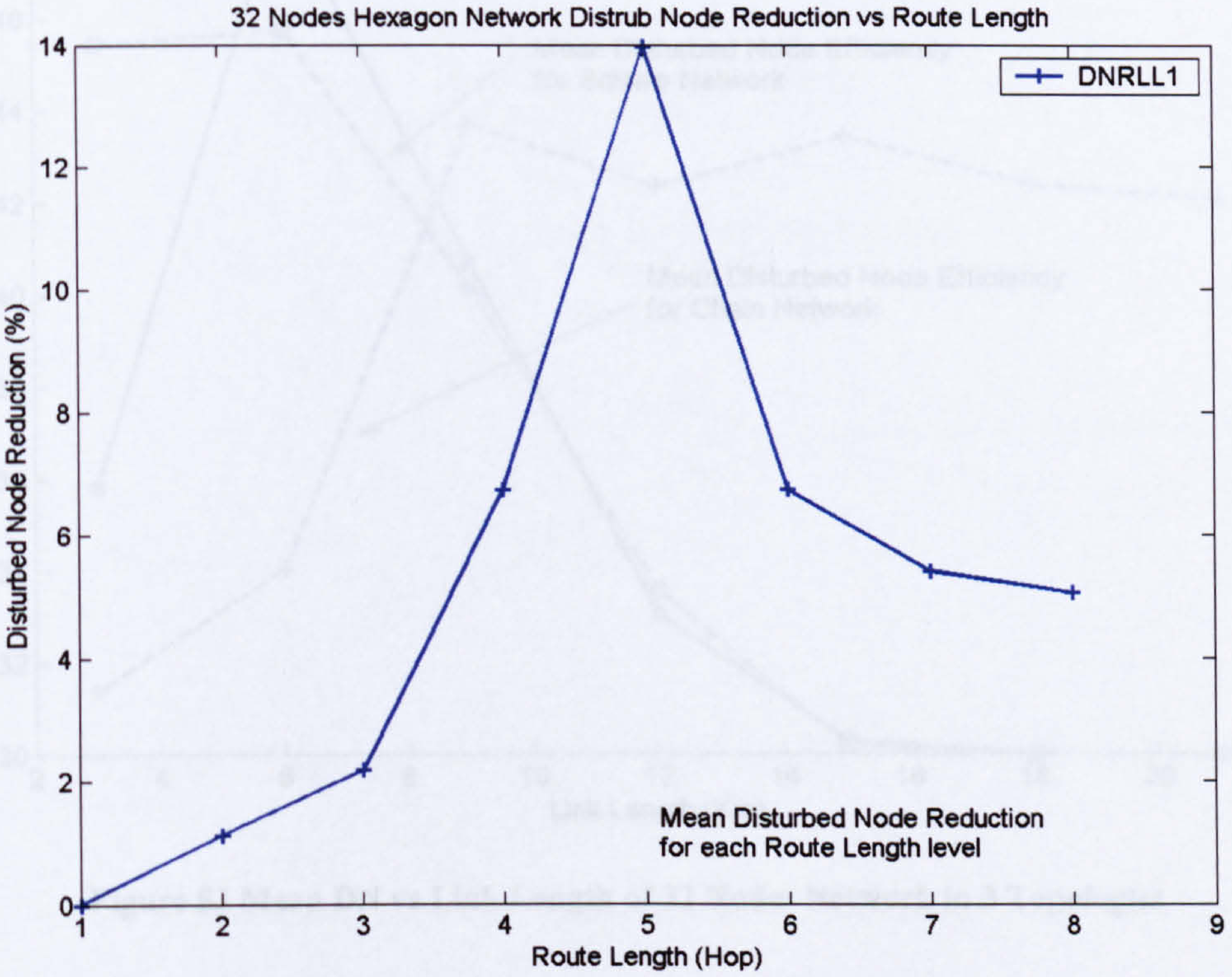


Figure 52 Disturbed Nodes Reduction of MIR in 32 Nodes Hexagon Network

The percentage reduction in the number of DNs arising from use of the MIR algorithm for a fixed route length (measured in hops), is shown for a hexagonal topology in Figure 52. This graph shows that the biggest improvement takes place at a medium multi-hop route length. In the simulation, MIR provided an average reduction of up to 14% in DNs in comparison with the SP which has been achieved. The highest DNR arises for such route lengths because changes in the route give the maximum benefit. In the case of the longer routes both, MIR and SP will disturb nodes multiple times on successive hop transmissions, which means that the reduction in the number of hops will have proportionally less effect.

The mean number of disturbed nodes can be examined for each topology as a whole for different transmission link lengths. These results highlight the effect of the different topologies – see Figure 53. MIR performs differently in the three different topologies.

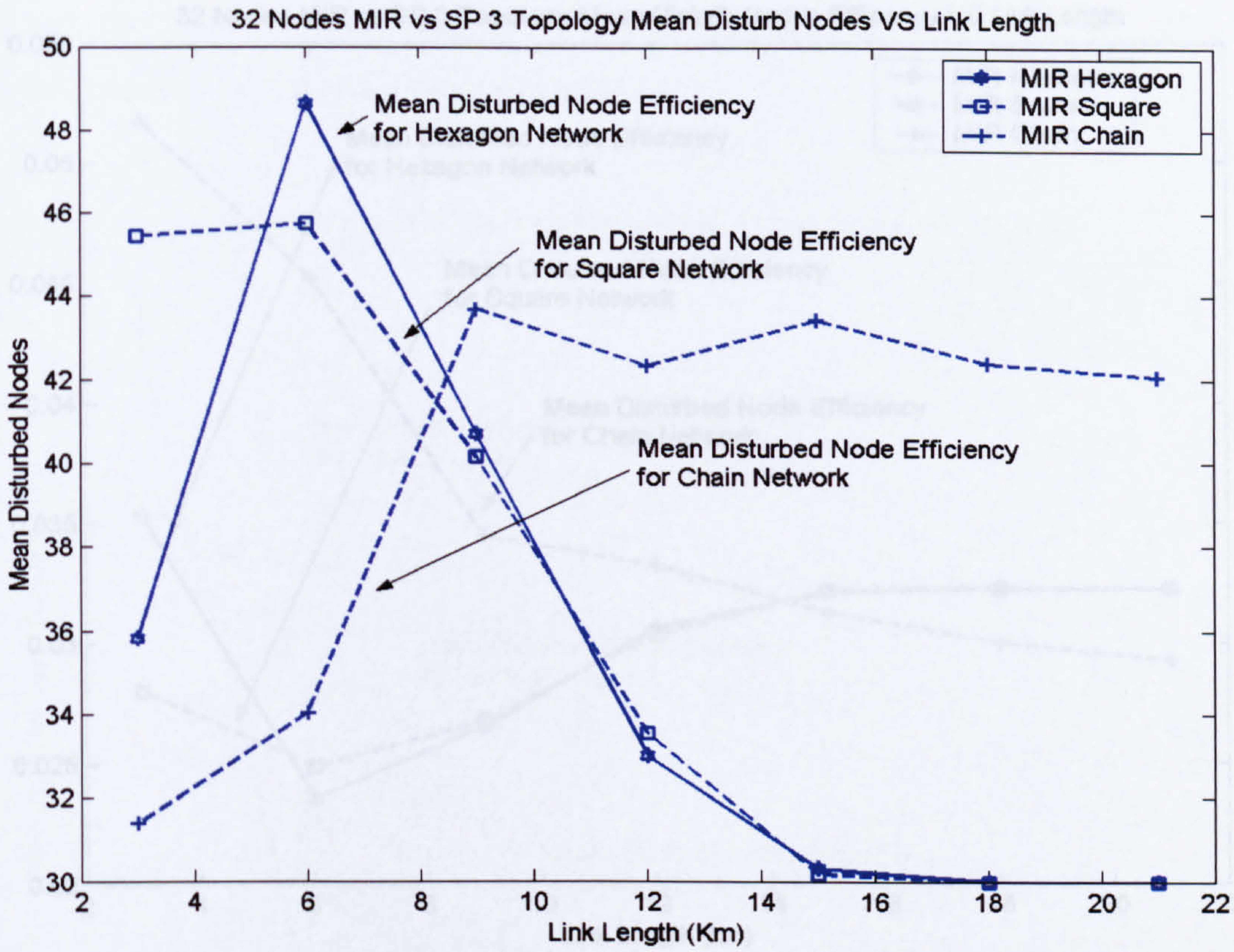


Figure 53 Mean DN vs Link Length of 32 Nodes Network in 3 Topologies

In the hexagon and square networks, MIR has a lower number of DNs when transmission link length is very short and very long, but disturbs more nodes at a medium link length. The minimum link length results in a sub-optimum hop distance that means that nodes are repeatedly disturbed for little advancement along the link.

With a chain network the effect is more pronounced, MIR performs well for short link lengths, and remains high for all link lengths above 8 km. This is because there is no alternative choice of route and the number of disturbed nodes is similar. If the graph were extended to a link length of 32, one would expect the mean number of disturbed nodes to reduce to 30.

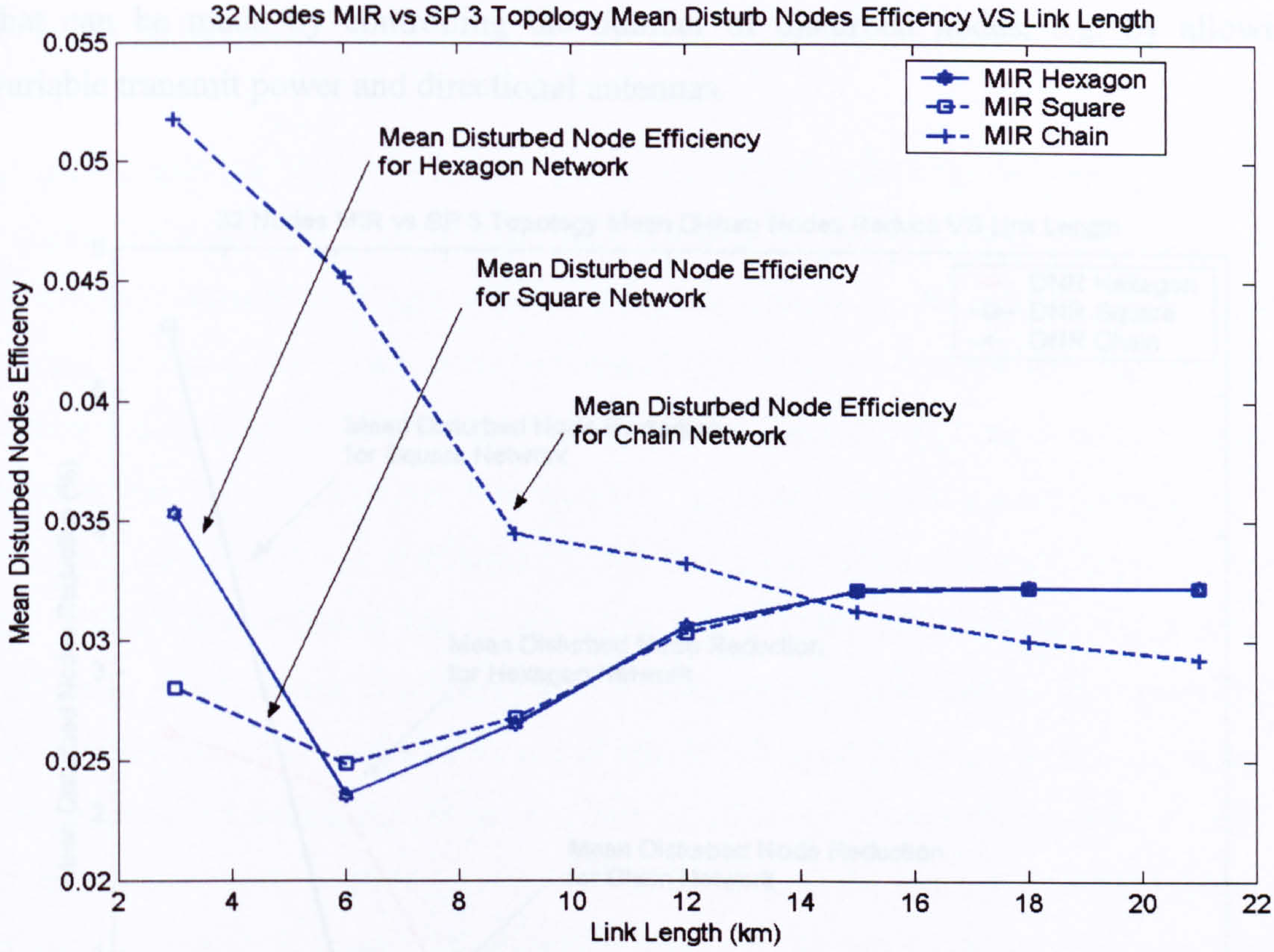


Figure 54 MIR Mean Disturbed Nodes Efficiency vs. Link Length in 3 Scenarios

Another performance measure is the Mean Value of DN Efficiency (MDNE), which indicates the fraction of nodes involved in the transmission compared with the optimum number. It is a measure that indicates how many percent of nodes were involved in a transmission compared with the optimum number that should be involved in the transmission (i.e. disturbing no nodes). The MDNE is calculated using the following formula:

$$MDNE = \frac{\textit{Optimum number of Receiving nodes}}{\textit{Actual number of Receiving nodes}} \qquad (6.3)$$

Figure 54 shows the MIR performance measured in MDNE for three network topologies. The chain network has by far the best efficiency, since the average number of DNs is the lowest. Hexagon and square topologies have lower MDNE, since their average DNs are higher. This is largely due to the node connectivity. If a node has good connectivity, it is likely to disturb more nodes – i.e. fewer simultaneous transmissions can take place. The low values of efficiency indicate that there is much improvement that can be made by controlling the number of disturbed nodes, e.g. by allowing variable transmit power and directional antennas.

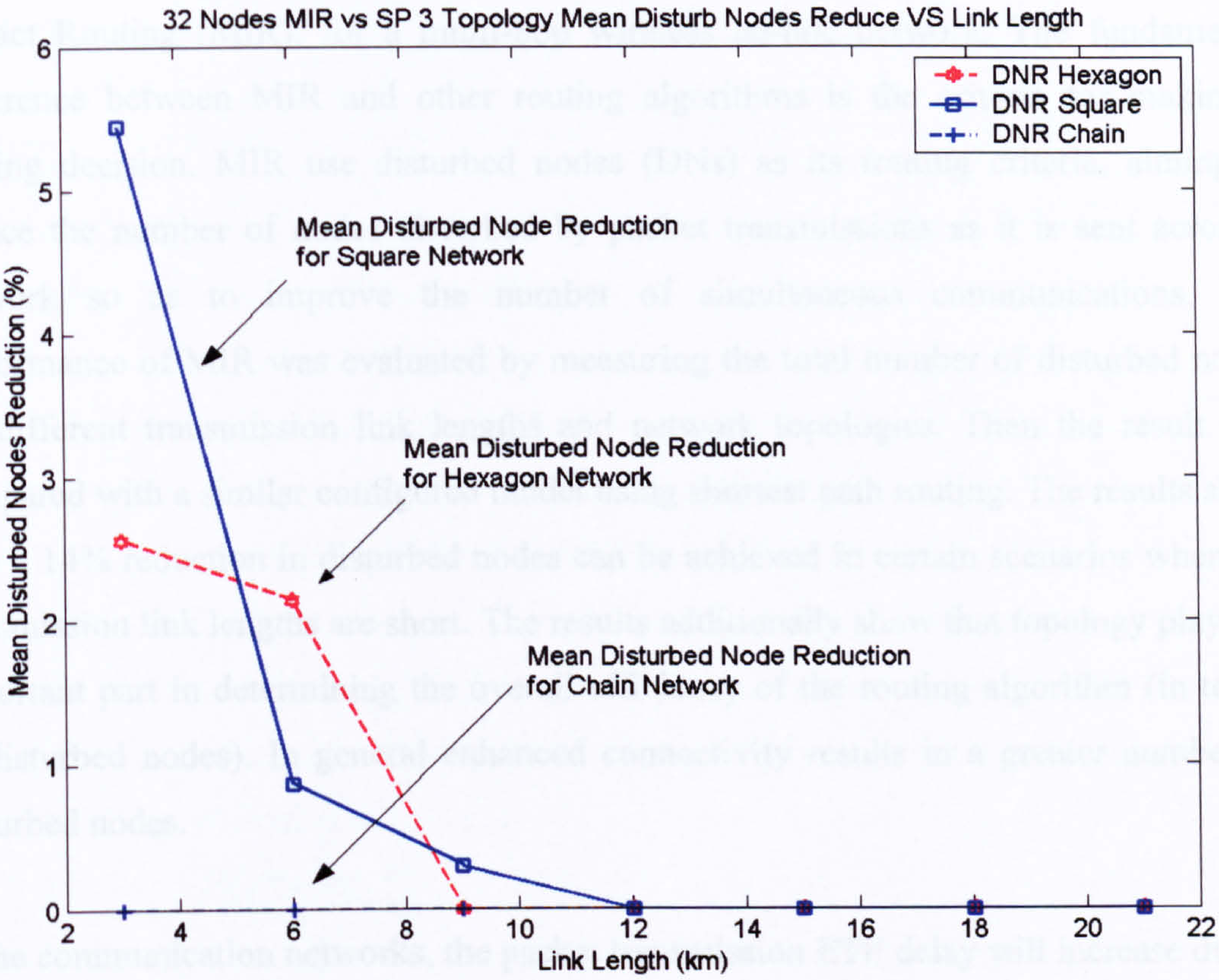


Figure 55 Mean DNR vs. Link Length of 32 Nodes Network in 3 Topologies

Figure 55 shows the effect on disturbed node reduction on each topology for different link lengths and it can be seen that MIR benefits the square topology best. This is because the topology requires many more hops to get across the network compared with the hexagon, which coupled with the reasonably flexible route choice ensures the biggest benefit.

6.4 Conclusion

In this chapter, a new interference-aware routing scheme was discussed, Minimum Impact Routing (MIR), for a multi-hop wireless ad-hoc network. The fundamental difference between MIR and other routing algorithms is the criteria for making a routing decision. MIR use disturbed nodes (DNs) as its routing criteria, aiming to reduce the number of nodes disturbed by packet transmissions as it is sent across a network so as to improve the number of simultaneous communications. The performance of MIR was evaluated by measuring the total number of disturbed nodes for different transmission link lengths and network topologies. Then the result was compared with a similar configured model using shortest path routing. The results show how a 14% reduction in disturbed nodes can be achieved in certain scenarios when the transmission link lengths are short. The results additionally show that topology plays an important part in determining the overall efficiency of the routing algorithm (in terms of disturbed nodes). In general enhanced connectivity results in a greater number of disturbed nodes.

In the communication networks, the packet transmission ETE delay will increase due to many causes, e.g. inefficient routing generates additional control traffic or packet retransmission. This results in the relaying burden that generates redundant traffic load dominating communication links, reducing the network capacity and throughput. MIR measures interference impact using DNs, and then compares this measure with an optimal node density threshold, maximum DNs (DN_{max}). When the number of surrounding nodes increases, then the transmitting node using DIIP to measure local DNs (DN_L) reflects this density, and compares it with a preset DN_{max} threshold. If the DN_L is greater than DN_T then two changes will be made. First, the transmit power level will decrease to a matching level. Then the working node will recalculate and determine its LCG members within the new radiation range. When the opposite situation happens, DN_L decreases, the working node will increase its transmitting power, so that its DN_L level remains within the limit of a DN_T.

Chapter 7 The MIR-VA Scheme

Contents

7.1 Introduction

MIR with Variable-transmit-power plus Asymmetrical-routing (MIR-VA) is the further improved version of MIR. The routing scheme introduced in the last chapter can only choose a particular Local Communication Group (LCG) size according to a fixed identical transmit power in each node. The result of the previous simulations show that the model using MIR reduces the number of Disturbed Nodes (DNs) as predicted, and consequently increases the concurrent transmissions in the network, hence improving the system and user capacity. MIR-VA, using the same total DN measure along a path as its routing decision making criterion, and will select a path with the minimum number of DNs for transmission or relay.

However, the fixed transmit power in each node does not reflect the real network behaviour, where every node could send their signals at a different strength to achieve adequate SINR for ensuring link quality. In addition, the interference impact measurement is fixed for every node, which means there is no coordination between nodes, and the node density around a transmitting node is not reflected by the transmitting node's LCG. In this case, the pure MIR scheme used by the simulation can only find a better configuration, in terms of reducing DNs, by manually adjusting the previously identical transmitting power in each node.

In order to simulate interactive operations of MIR nodes, a variable transmit power feature is added into the scheme. Consequently, power variation triggers the link length variation, which generates asymmetrical links in the network, and in the size of an LCG, and can be adjusted to reflect the neighbouring node density of a transmitting node. Therefore the routing operation of pure MIR also incorporates an asymmetrical routing strategy for adaptation to the asymmetrical link environment.

This chapter will introduce, the theoretical characteristics of MIR-VA, in which concerns Variable Transmit Power (VTP), asymmetrical routing strategy, spatial reuse theory, consequently the Time Sequenced Interference Region (TSIR) conceptual model, the derivation of system and user capacity based on the measure of interference impact, and finally the MIR-VA protocol in section 7.2. Section 7.3 looks at the performance of the simulation model using MIR-VA algorithm.

7.2 MIR-VA Theoretical Characteristics

In addition to the Distributed Interference Impact Probing (DIIP) mechanism, and the LCG concept, the MIR-VA integrates and incorporates the VTP and the asymmetrical routing scheme into the pure MIR algorithm.

7.2.1 Variable Transmit Power

The transmit power control has been used in conventional wireless communication networks, such as cellular networks, for energy conservation and improving the likelihood of good link quality, and increasing capacity, so that each node can communicate with a base station with minimum transmit power, yet with sufficient signal-to-interference ratio on the wireless link (referred to as short for link quality in the rest of this thesis) [5, 13, 42].

In cellular telecommunication systems cellular phones adaptively adjust their power level so as not to swamp all the other users in the system [5, 13, 42]. Cell phones close to a base station/cell tower transmit using very little power. As they get further away, the base station/cell tower tells them to transmit using more power. This adaptive transmission has two benefits [5, 13, 42]:

- A fixed transmit power level would either (close to the tower) use more energy than necessary, reducing the battery life of the cell phone, or transmit too quietly to be heard, reducing the range of the cell phone.
- The cell tower adjusts the power level of each cell phone so that at the *tower* they all have close to equal power levels. This makes it much easier to separate all the signals from each other; if one phone were much "louder" than the others, it would be more likely to bleed through into the other signals.

Power-awareness is one of the popular research fields of WANET. The purpose of power-awareness for WANET is summarized as follows:

- Energy conservation - increases energy consumption efficiency, to enable devices to operate new functions, and prolonging battery lifetime for a single node and the whole network. This ensures that smaller, lighter, and more environmentally friendly power sources can be used
- Interference Control - adjust transmit power level to reach adequate link quality without prohibiting all other nodes from communicating. Hence this increases the spatial reuse ratio in WANET, and increases user capacity.

Conventional routing algorithms determine the most appropriate routes through the network using a route selection constraint. Usually this is achieved by using criteria such as the shortest path length (either physical distance or minimum number of hops), minimum energy consumption, or selecting nodes with the longest battery life [3, 7].

The latter two requires cross-layer knowledge to be used, especially as they will often adjust the transmit power [100].

A wide number of constraints can be used to select routes that have the minimum interference impact on the environment. This can be particularly helpful in congested environments, possibly where multiple networks are sharing a common band (either using the same standard or differing standards), as well as helping to conserve battery power. Impact is measured by determining the number of nodes that are disturbed by a transmission, or the overall levels of interference on the channel as shown in Figure 58. If the interference level is high, then it is possible that transmissions on the same channel will excessively interfere with other nodes [100].

MIR uses the DN measure as the routing criterion, which can be used in a number of different ways. Disturbed nodes may be nodes that can actually receive unwanted packets, or alternatively, just give rise to interference, reducing the SINR. Fixed transmit power routes which disturb fewest nodes can be selected [100].

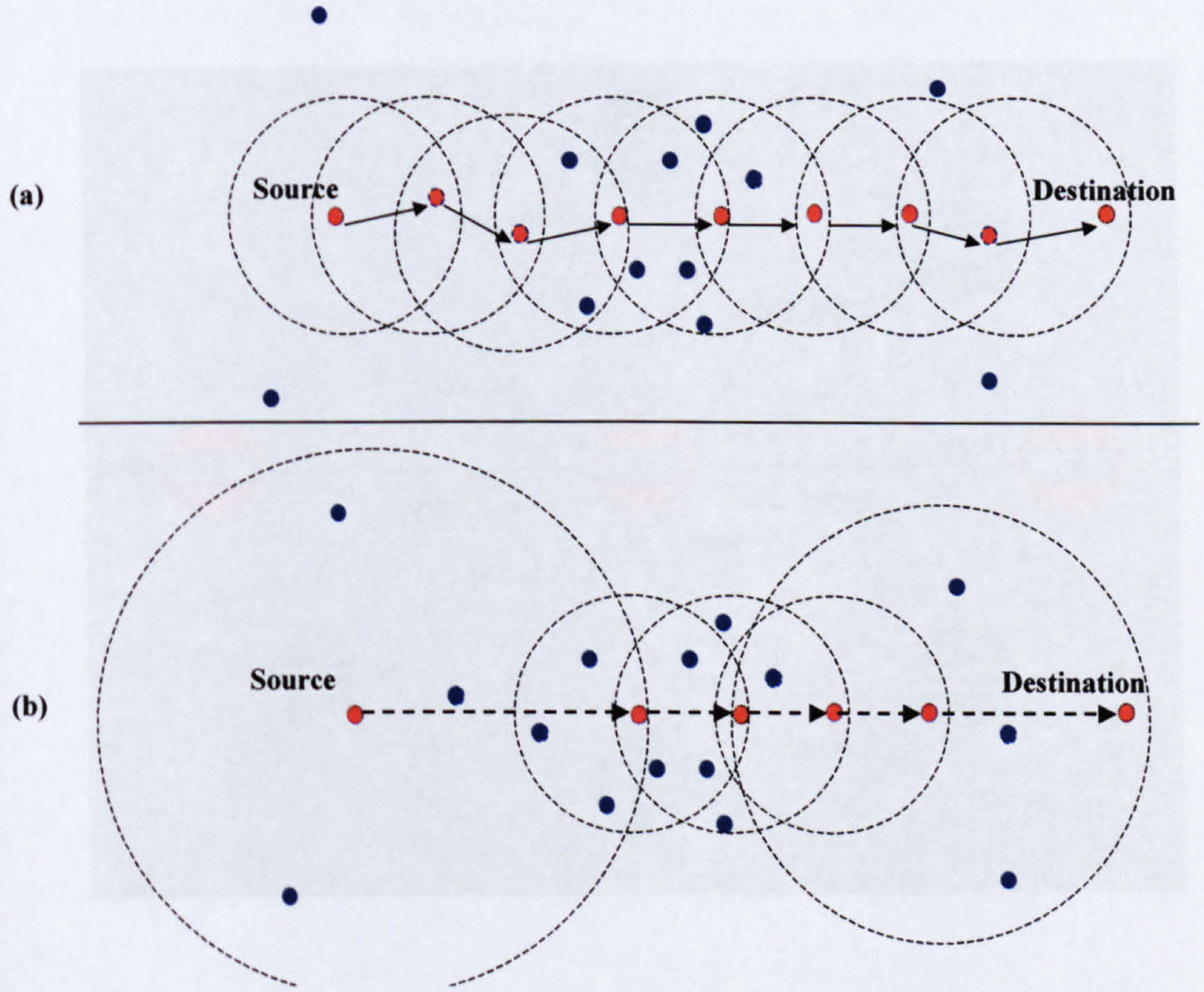


Figure 56 benefits of variable transmit power and minimum impact routes

(a) Pure MIR (b) MIR with VTP

The nature of WANET has given power control a new meaning, which is to improve spatial reuse ratio in the network. MIR-VA incorporates VTP into the pure MIR, taking into account measures of local interference impact, and adjusts its transmit power to dynamically form its own LCG, as shown in Figure 59. Such a scenario shows fixed-transmit-power routing, Figure 59 (a), has a greater number of hops, resulting in multihop relays, hence longer packet relaying delays. If variable transmit power is used, Figure 59 (b), the transmit power can be varied to disturb a set number of nodes. In this way, high transmit powers can be used in less congested parts of the network, resulting in fewer hops, with lower transmit powers used on parts of the route that are in close proximity to other networks. This provides more of a balance between disturbance and packet delay.

7.2.2 Asymmetrical Routing

As discussed earlier in chapter 5, the variable transmit power results in asymmetrical links on both directions of a path, hence the asymmetrical routing strategy has been developed especially to resolve this problem.

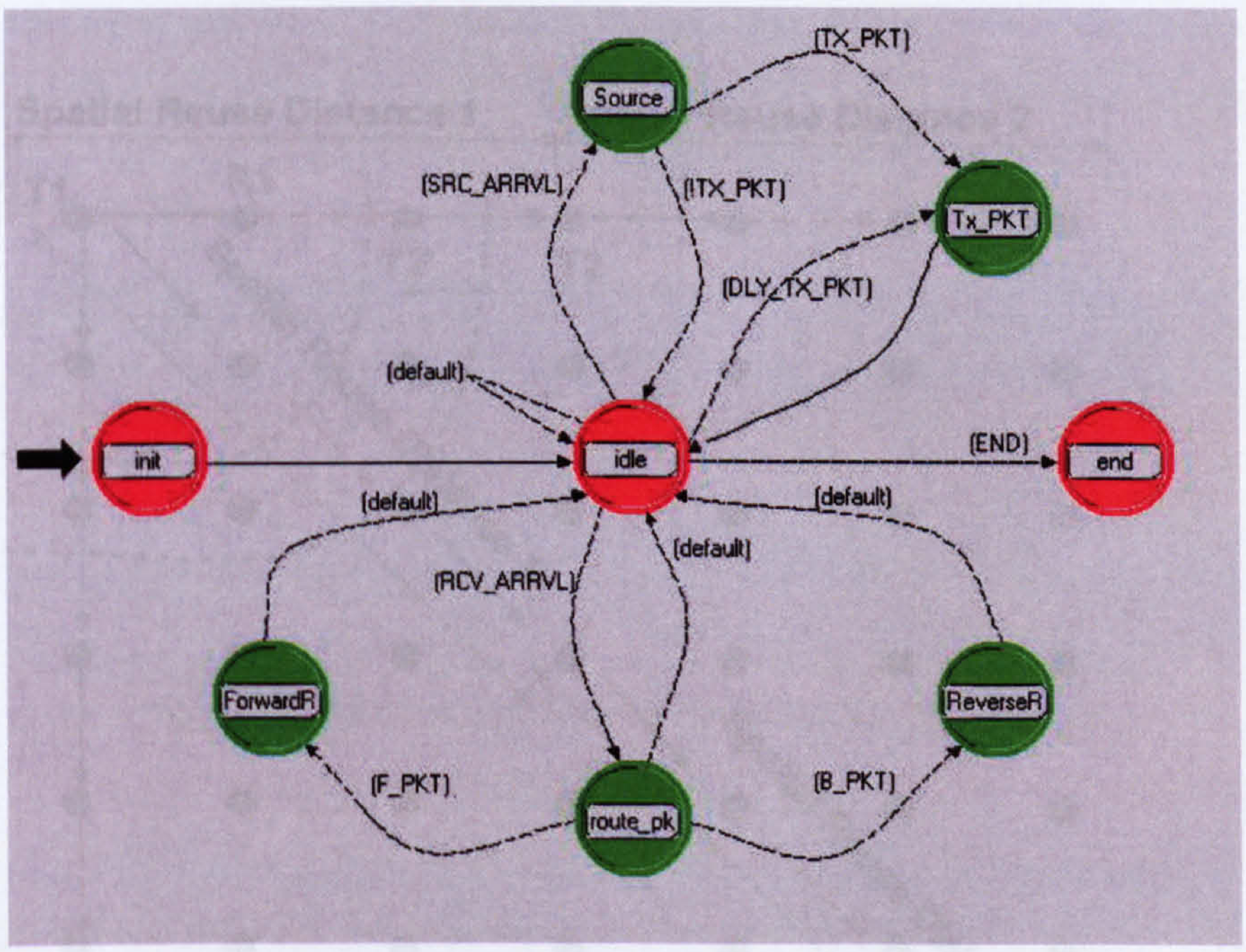


Figure 57 Hub Processing model for Asymmetrical Routing

The lower three state symbol in Figure 57 is the upgraded hub process model that splits the original "route_pk" state into three, so that the model is able to process the forward and reverse routing information separately.

7.2.3 Spatial Reuse and Carrier Sensing Multiple Access

The concurrent transmission, in systems that use Carrier Sensing Multiple Access (CSMA) for collision control, has a safe Spatial Reuse Distance (SRD) problem. This means in a shared spectrum environment, where it is assumed that hop length and node density are constant, and all the transmitting nodes are co-channel nodes, when one node is transmitting, the concurrent transmission that will not cause collisions must take place outside of a safe SRD distance (e.g. two hops between concurrent transmitting nodes, as the uniformed network shown in Figure 58).

For instance in Figure 58, two concurrent co-channel transmitting nodes, T1 and T2, have a safe distance to transmit simultaneously. In which case T1's receivers (i.e. R1 and other receivers inside T1's transmission radius) will not receive potential interruptions from nodes that are inside of the ring region between T1 and T2.

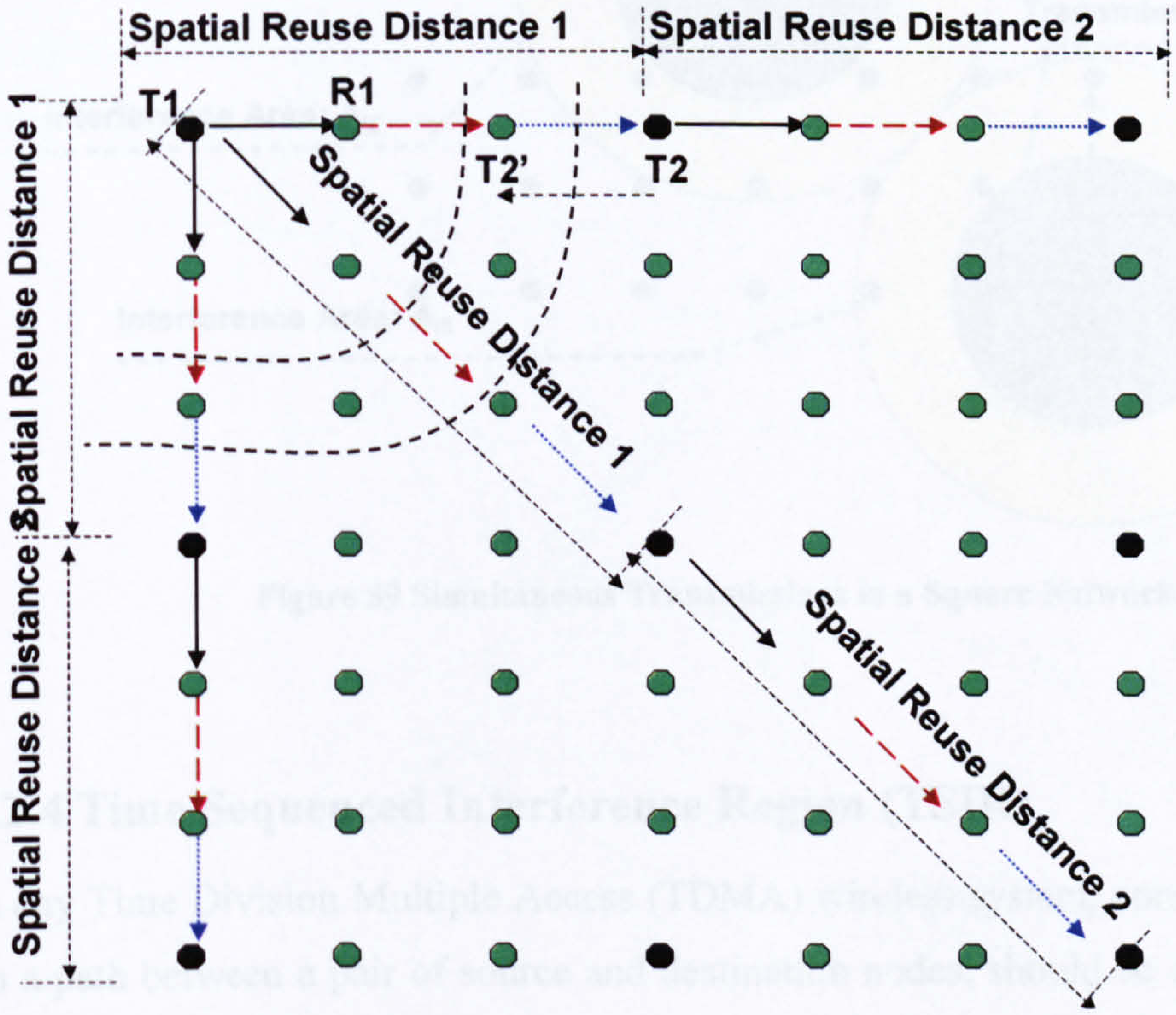


Figure 58 Safe Spatial Reuse Distance in a regular network

Otherwise, if T2 moves to the location T2', and transmits concurrently, the hidden terminal problem is likely to occur, and simultaneous transmissions of T1 and T2' could collide.

Figure 59 shows a simultaneous transmission scenario in a unified square network. This theory inspired later spatial reuse theory derived for MIR to calculate system and user capacity using the number of disturbed nodes. From this regular matrix of nodes, the spatial reuse ratio is derived. More detail about the capacity derivation is shown in section 7.2.5.

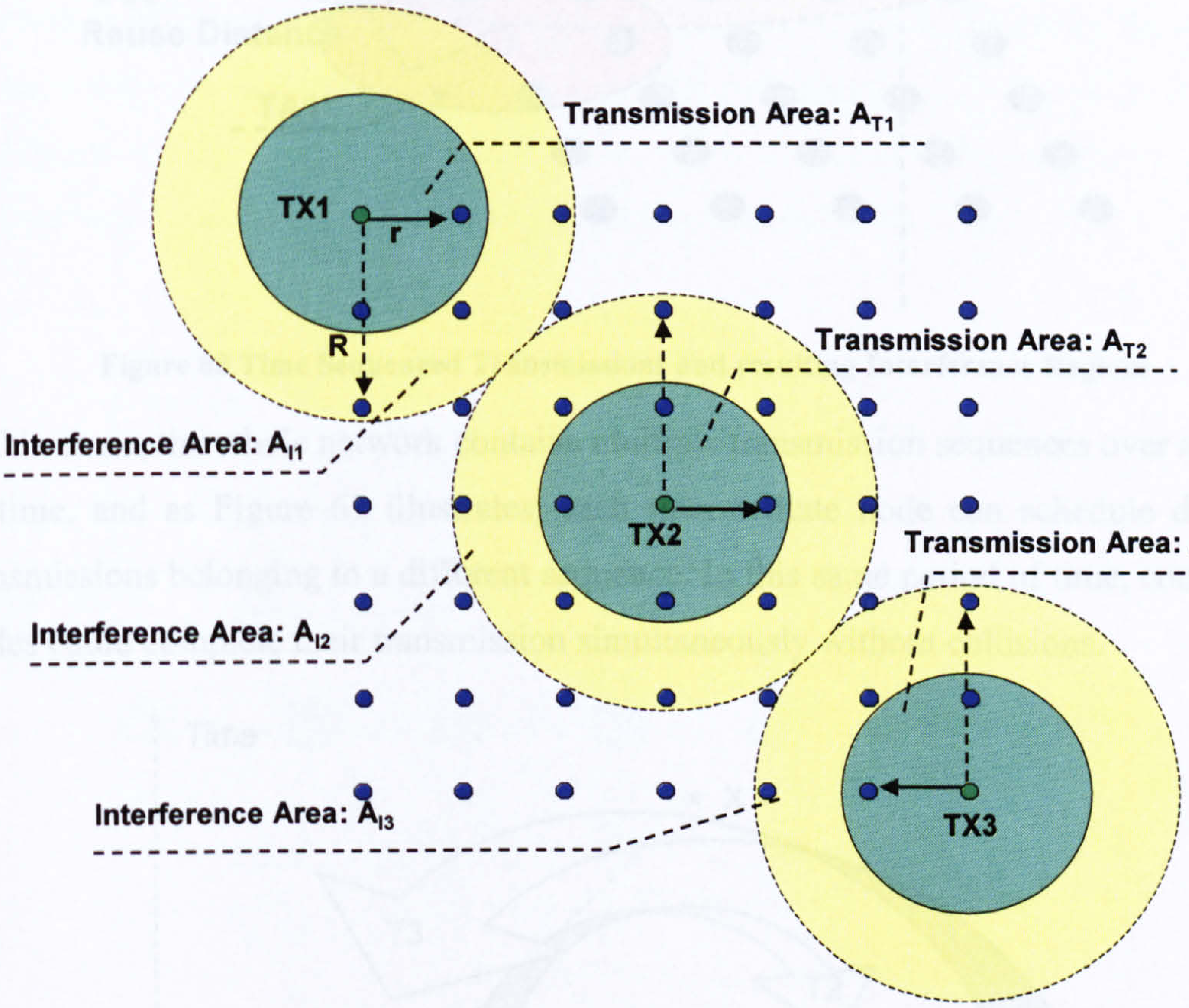


Figure 59 Simultaneous Transmissions in a Square Network.

7.2.4 Time Sequenced Interference Region (TSIR)

In any Time Division Multiple Access (TDMA) wireless system, considering all nodes on a path between a pair of source and destination nodes, should be considered. These therefore generate a series of transmissions as well as interference footprints along the path. The overall interference region along this path is referred to as the Time Sequenced Interference Region (TSIR), as shown in Figure 60.

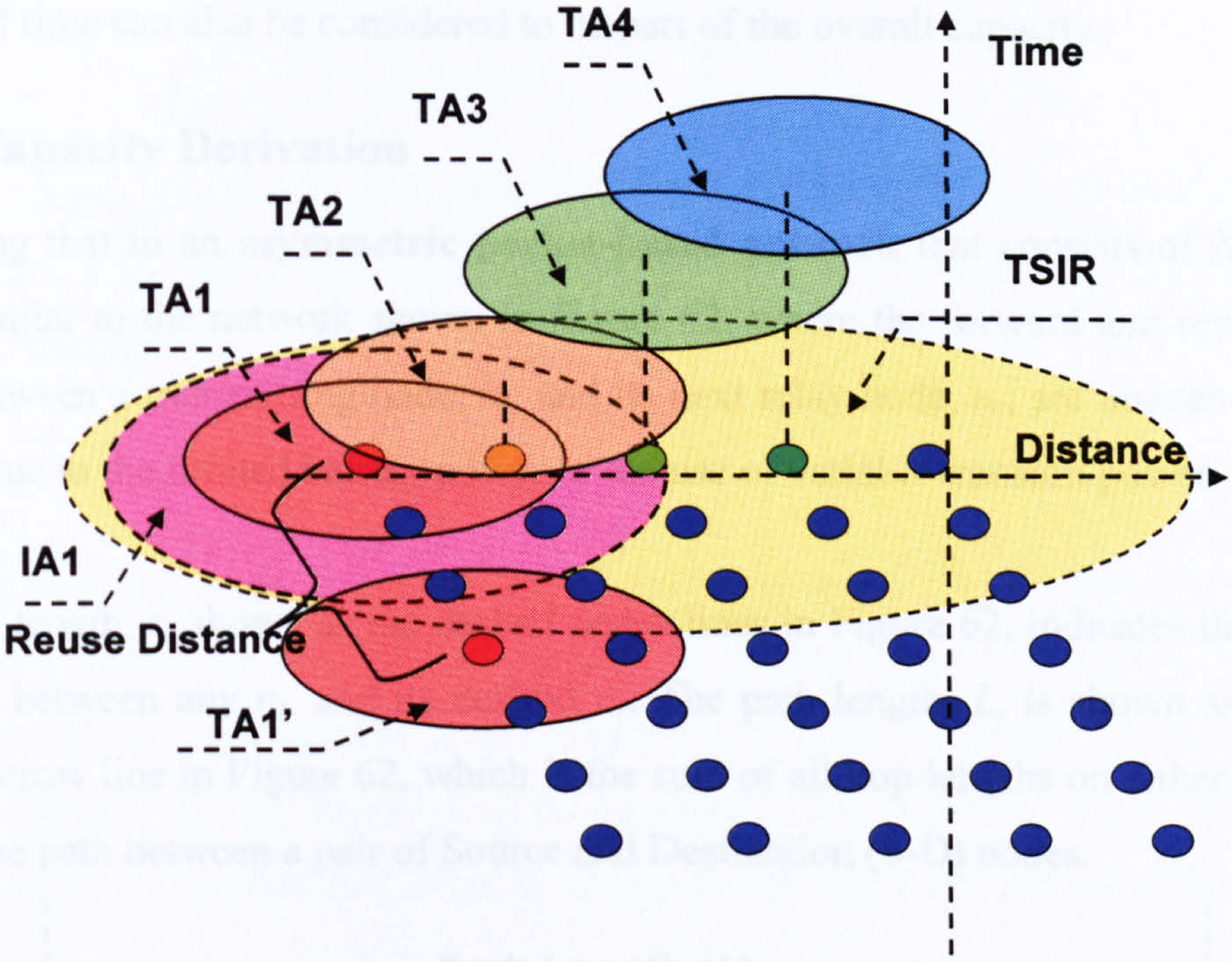


Figure 60 Time Sequenced Transmissions and resulting Interference Regions

In this sense, the whole network contains multiple transmission sequences over a period of time, and as Figure 61 illustrates, each intermediate node can schedule different transmissions belonging to a different sequence. In this same period of time, concurrent nodes could complete their transmission simultaneously without collisions.

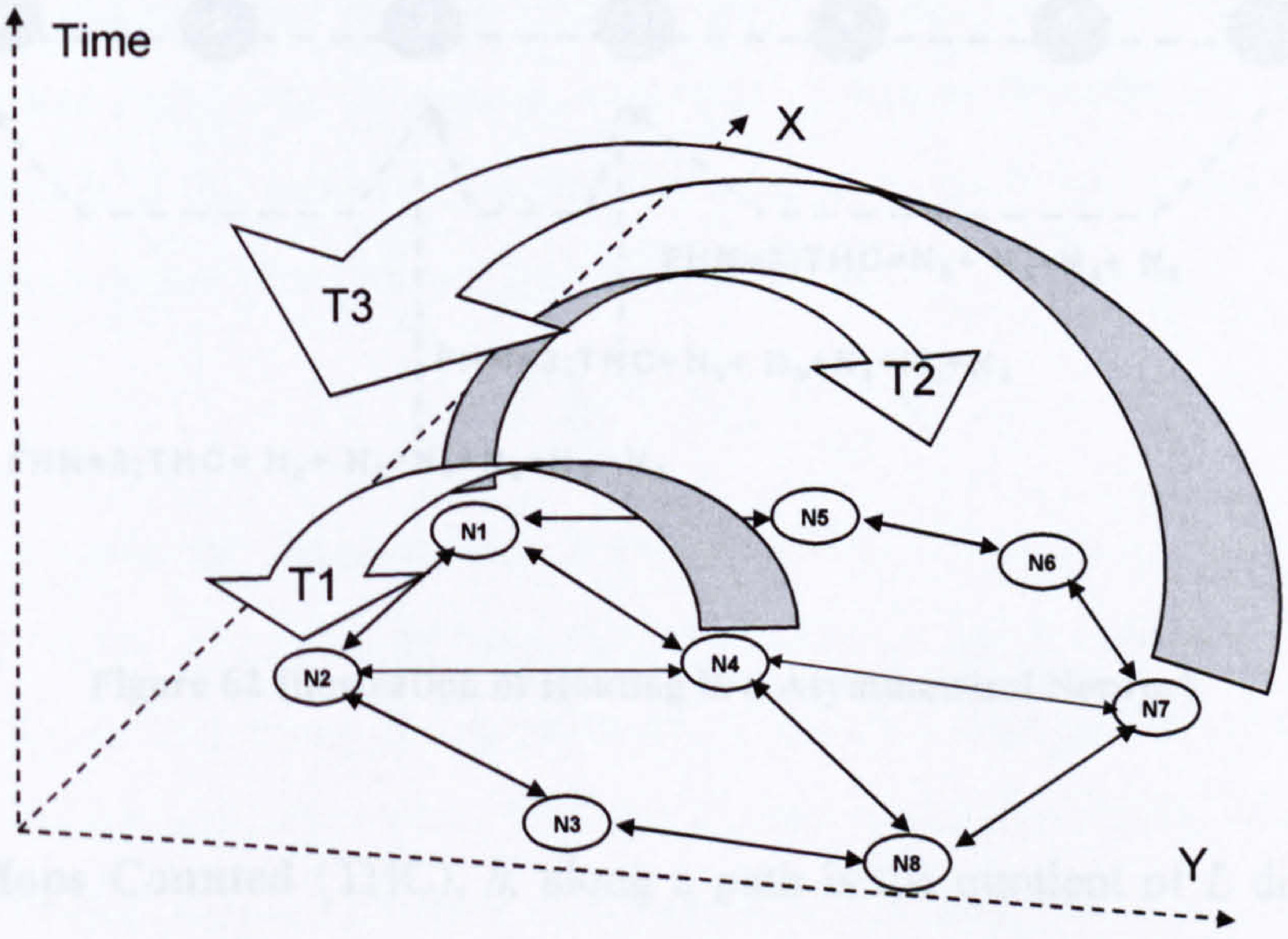


Figure 61 Time Sequenced Transmission Groups Along Different Path

If we consider all the nodes that receive transmission impacts within a TSIR is a subnet or a part of the whole network, then the capacity of such regions existing over a short period of time can also be considered to be part of the overall capacity.

7.2.5 Capacity Derivation

Assuming that in an **asymmetric packet-based network** that consists of N nodes in total, similar to the network shown in Figure 62, where the forward and reverse radio links between a transmitting node, n_t, and its next relay node, n_r, are uneven in length. This is due to the limited battery power or the use of variable transmit power.

The hop length, r, shown as the dashed arrow lines in Figure 62, indicates the one-hop distance between any n_t, and its desired n_r. The path length, L, is shown as the solid double arrow line in Figure 62, which is the sum of all hop lengths on either a forward or reverse path between a pair of Source and Destination (S-D) nodes.

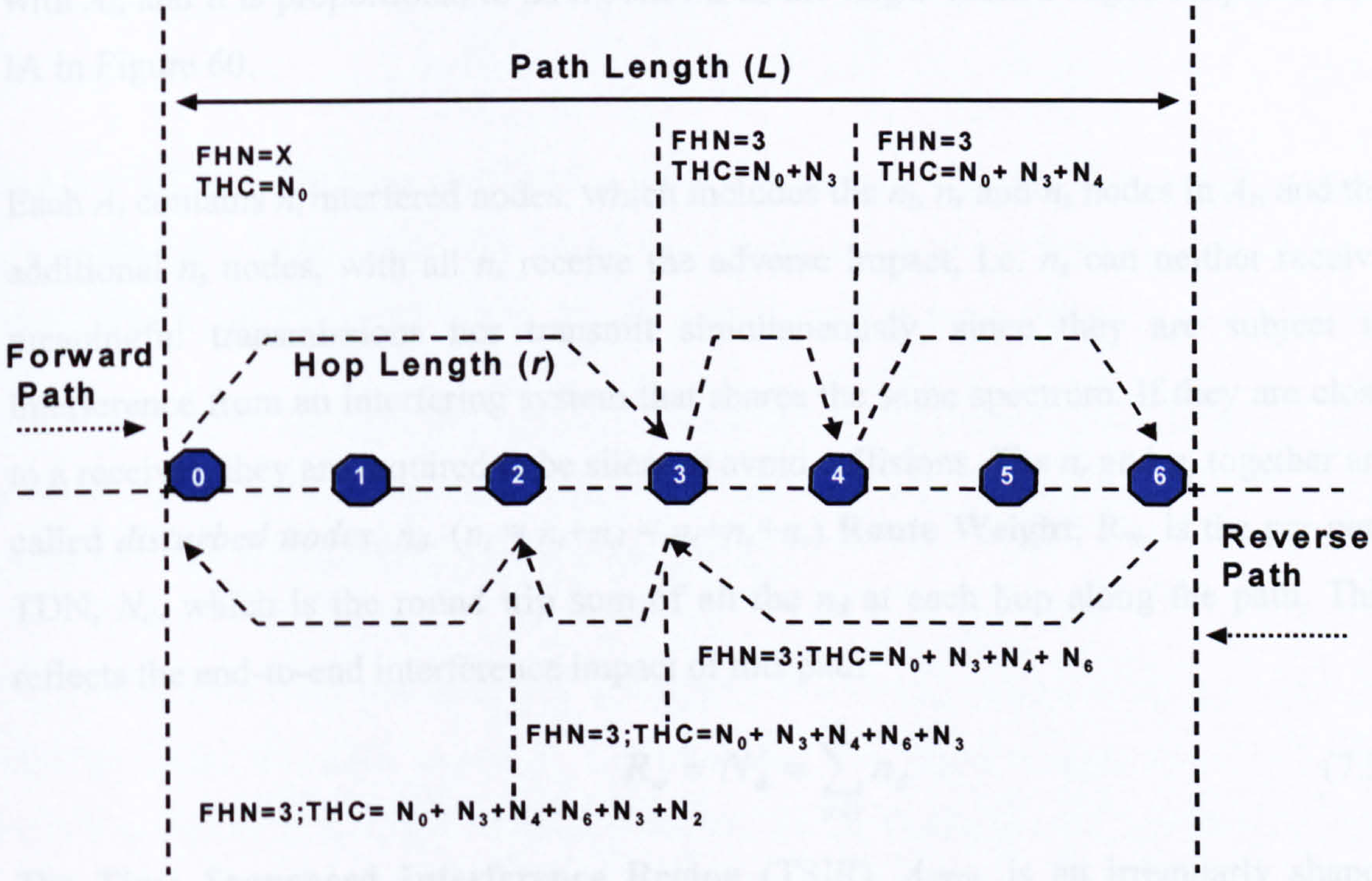


Figure 62 Illustration of Routing in a Asymmetrical Network

The **Total Hops Counted** (THC), h, along a path is the quotient of L divided by the average r. This measure can also represent the total number of relays that is equal to the number of transmitting nodes per-path, N_t, which reflects the number of end-to-end relaying on a path, as shown in equation 7.1.

$$\boldsymbol{h = \frac{L}{r} = N_t = \sum_{t=1}^{h} n_t} \tag{7.1}$$

The hop length, r, is inversely proportional to the maximum transmit power, P_m, at each transmitting node. The **power consumption gain**, P_c, for each hop is the quotient of P_m divide r (watt/km).

$$\boldsymbol{P_c = \frac{P_m}{r}} \tag{7.2}$$

Each *transmitting node*, n_t, creates two interfering areas. The first is a ***transmission area***, A_t, with n_t at its centre with a radius of r, shown as disk areas, TA, as shown in Figure 60. Inside each A_t there is one transmit and receive node pair (n_t, n_r) surrounded by ***silent nodes***, n_s. The second is an ***interference area***, A_i, which shares the same core with A_t, and it is proportional to an A_t, shown as the larger dashed edged elliptical area IA in Figure 60.

Each A_i contains n_i interfered nodes, which includes the n_t, n_r and n_s nodes in A_t, and the additional n_s nodes, with all n_s receive the adverse impact, i.e. n_s can neither receive meaningful transmissions nor transmit simultaneously, since they are subject to interference from an interfering system that shares the same spectrum. If they are close to a receiver, they are required to be silent to avoid collisions. The n_r and n_s together are called ***disturbed nodes***, n_d. ($n_i = n_t+n_d = n_t+n_r+n_s$) **Route Weight**, R_w, is the per-path TDN, N_d, which is the round trip sum of all the n_d at each hop along the path. This reflects the end-to-end interference impact of this path.

$$\boldsymbol{R_w = N_d = \sum_{d=1}^{h} n_d} \tag{7.3}$$

The **Time Sequenced Interference Region** (TSIR), A_{TSIR}, is an irregularly shaped interference region along a multi-hop path, shown as the largest elliptical area in Figure 60. A_{TSIR} consists of a series of footprints of A_i, which are generated by all nodes n_t on this path in a time-sequence. Each S-D pair creates a per-path TSIR, which consists of N_i interfered nodes that are the sum of all the n_i on this path ($0< n_i = n_t + n_d \leq N_i$; $N_i \subset N$). Hence we derive the following:

$$N_l = \sum_{l=1}^{h} n_l = \sum_{t=1}^{h} n_t + \sum_{d=1}^{h} n_d = N_t + N_d \tag{7.4}$$

Any TSIR is independent from other TSIRs created by different S-D pairs, since transmissions can be scheduled in turn to ensure successful reception. Any node can be a member of more than one TSIR in different time periods. If we assume the network size and node density is constant, and each node can transmit over a one hop distance, but interfere with its two hop neighbours, then A_l is greater than A_t, but less than or equal to a A_{TSIR} or the whole network A_{net}. ($0 < A_t \le A_l \le A_{TSIR} \le A_{net}$; $A_t \subset A_l \subset A_{TSIR} \subset A_{net}$)

7.2.5.1 Spatial Reuse and Capacity Calculation

The per-path **spatial reuse number,** s_p, and the per-network **spatial reuse number,** s_n, indicate the number of spatially divided simultaneous transmission areas, TA', as shown in Figure 60, of another TSIR and within the whole network respectively.

$$s_p = \frac{N_l}{n_l} = \frac{N_l}{n_t + n_d} \quad ; \quad s_n = \frac{N}{N_l} = \frac{N}{N_t + N_d} \tag{7.5}$$

Gupta & Kumar [20] have demonstrated that if the node's location and traffic patterns are optimally chosen, then the *throughput capacity* derived from a geographical analysis of a *arbitrary network* is $\frac{W}{\sqrt{N}}$, N is the total number of nodes in the network, and W is the channel capacity. Toumpis & Goldsmith [21] show that in a time-division scheduling asymptotical network, the *network capacity* is the maximum achievable rate of the source and destination transmission schemes.

Combining these two concepts, and developing a new way of describing the network capacity based on **path capacity,** C_p, calculates the regional capacity of a TSIR including factors affecting the routing decision. These factors are the *interference and relay impact* measured using N_l in equation 7.1 reflecting the THC, and the N_d in equation 7.3, reflecting the **Total Disturbed Nodes** (TDN); the path spatial reuse number, s_p, and the power consumption gain, P_c, in equation 7.2.

Consider that each TSIR is an independent sub-network consisting of N_l nodes over a period of time, and assume that the maximum channel capacity W is 1 Erlang, then adapting Gupta & Kumar's capacity bound, the path capacity C_p can be derived as shown below in equation 7.7.

$$C_p = \frac{1}{\sqrt{N_l}} \times s_p \times P_c = \frac{1}{\sqrt{N_t + N_d}} \times \frac{N_l}{n_l} \times \frac{P_m}{r} \tag{7.7}$$

The ***network capacity***, C_n, is considered here to be the sum of all path capacities of several co-existing TSIRs over the round trip period. One node from each TSIR transmits concurrently and is sufficiently spatially divided from other transmissions in different TSIRs. In order to derive this new bound, we assume that these concurrent multiple transmissions will be continuous and will be kept spatially separate over the period of end-to-end transmission time.

$$C_n = \sum_{p=1}^{S} C_p = \sum_{p=1}^{S} \frac{1}{\sqrt{N_t + N_d}} \times s_p \times P_c \;\; ; \;\; C_n = C_p' \times s_n \tag{7.8}$$

The C_n can also be derived as the product of average path capacity, C_p', and the per-network spatial reuse number, s_n, as shown in equation 7.8. If N_d or N_t increases, the s_p will decrease, and consequently the path and network capacity will decrease.

7.2.6 MIR-VA protocol

In this improved version, MIR-VA inherits the previous foundations developed and published, but introduces asymmetric routing and a variable transmit power strategy. Initially when routing tables are empty, each node will broadcast packets initialised with routing information such as: First Relay Node (FRN), Next Relaying Node (NRN), Total Hops Counted (THC), Total Disturbed Nodes (TDN), lifetime, and sequence number etc.

By exchanging packets, routing information is disseminated to allow existing nodes and newly activated nodes to announce their existence, learn each other's identity and location, determine their immediate neighbours, and construct one entry for each known source in a local routing table etc. Each transmitting node n_t declares a LCG, i.e. a cluster of interfered nodes, n_i. Packets are broadcast within this group, and relayed by intermediate nodes on the forward and reverse paths separately on a hop-by-hop basis.

On a forward path as shown in Figure 62, the first relaying node's IDentification (ID) will be recorded in the packet's header as the FRN. Each intermediate relay node adds the local disturbed nodes on to its total DN, increasing its total HC, reducing its lifetime, and creating a new entry for this unknown source. When the packet successfully reaches its destination within its lifetime, its forward path routing information, together with a fresh lifetime and new sequence number, will be added onto a return packet back to the same source. The source address will be used as destination address for transmitting in a reverse direction.

On a reverse path, shown in Figure 62, intermediate nodes relay the packet, accumulate the total HC and the total DN based on the forward path measures, and reduce the packet's lifetime, until it returns to its original source node. The round trip total HC and total DN, i.e. forward and reverse path, and its FRN on the forward path will be recorded in this packet's header. If there was no record of this destination, or the previously recorded route weight in the entry of this destination is greater than the packet's route weight, the routing information in this packet's header will be used to update the routing table. Its FRN will be chosen as NRN in the local communication group for next transmission towards the same destination.

Intermediate nodes accomplish flood control by selecting and attaching an NRN's ID in each relaying packet according to its destination address. The node receiving this packet will forward it only if the node's ID matches the packet's NRN ID, or no NRN ID is attached for that destination. This minimizes duplicated packets. If the packet has a stale lifetime or sequence number, it will be discarded on arrival. Nodes using MIR select a route with the minimum route weight along a path. Route maintenance is achieved by updating the NRN's ID for each destination in each source node's routing table using the received packet.

Initially when routing tables are empty, every node starts to broadcast BPs, so that other nodes can learn of their existence, determine their position, and create entries in the routing tables. For example, if N_1 transmits a BP, all nodes within the LCG will receive the packet and fill in the N_1 row in the routing table. The highlighted column in Table 9, shown below, is the entry of N_1 in N_i's routing table (i does not equal 1). In this case, after a while every node will have a routing table consisting of one routing entry for each learned source node.

Identified Node ID (Source/Destination)	**Forwarding Node ID (Last/next)**	**Disturbed Nodes**	**Sequence Number**	**Other Routing Control factors**
N_1	N_{i-1}	20	015	…
N_2	N_{i-1}	15	203	…
…	…	…	…	…
N_{i-1}	N_i	3	251	…
N_i	N/A	N/A	N/A	…

Table 9 Example of MIR-VA routing table for Node N_i

When a node needs to transmit a data packet to a known source node, without knowing a predetermined route, the data packet will be flooded into the network in a hop-by-hop manner with a limited lifetime (measures by HC), and a sequence number. This ensures the data packet will not cause endless retransmissions and limits the number of forwarded packets. Every intermediate node that forwards this data packet will add its own ID, together with total number of local DNs into this data packet's header. By the time this data packet reaches its destination node, its recorded route weight, which is the total number of DNs along the path, will be compared with other data packets from the same source. The data packet with the recorded lowest route weight will be used to update the destination node's routing table with the last forwarding node.

When a node needs to send a packet in a reverse direction. The recorded 'source node ID' will be used as a new 'destination node ID', and the previously recorded 'last forwarding node ID' will become the next 'next forwarding ID'. In this way every node will soon determine a NFN for each identified node.

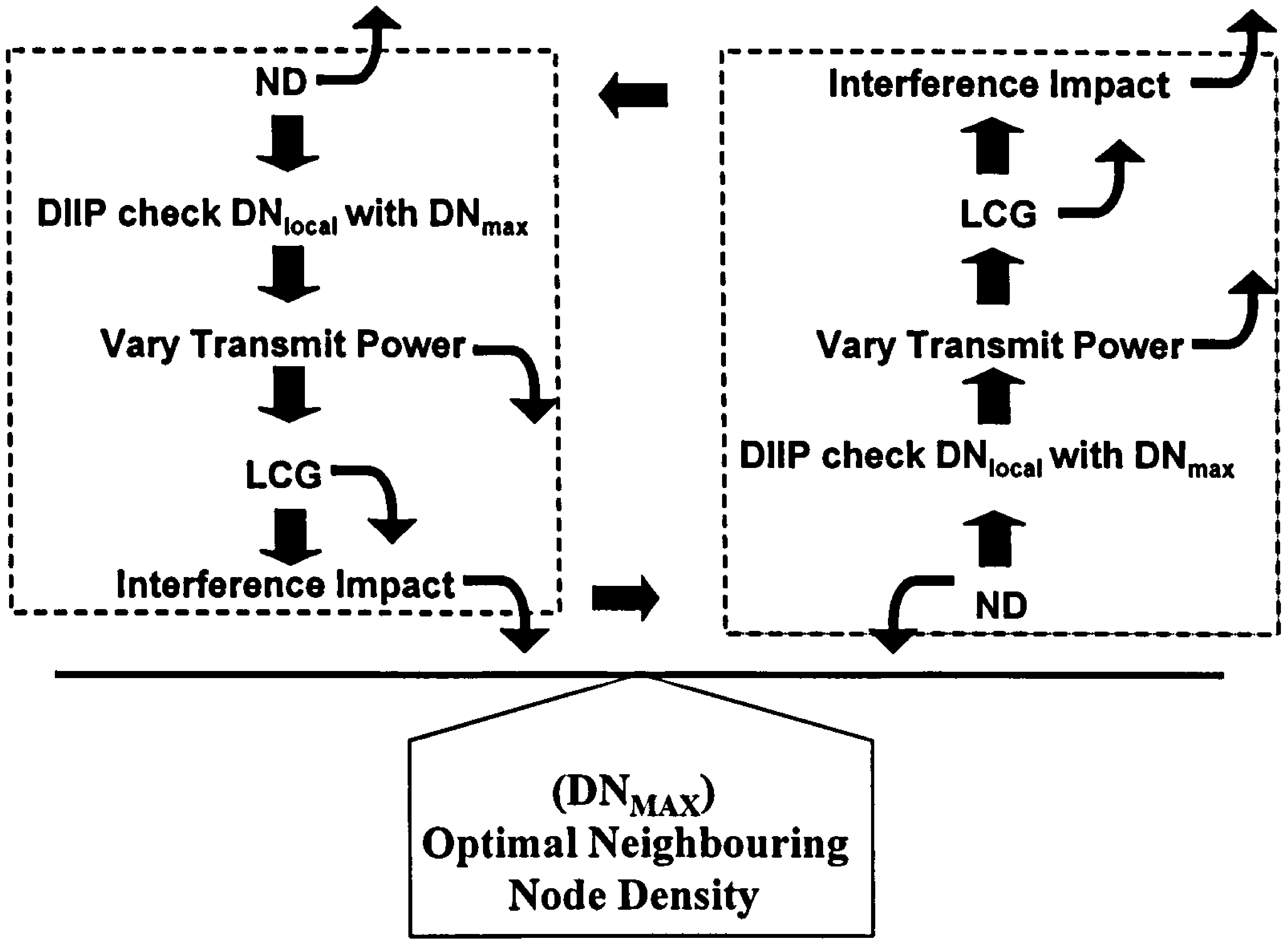


Figure 63 Balancing Interference Impact use DNmax threshold in MIR-VA

Route maintenance is achieved by updating the 'forwarding node's ID' by reading received data packets. Each source node only needs to forward a data packet to its next forwarding node within the LCG, which is one-hop away. Packets are only forwarded if the 'next forwarding node' indicates that it is the current node's ID, or no route is known to that destination. This stops data packets flooding the network, and minimises the number of duplicated packets. A data packet will time out after its lifetime limit has run out. If any changes take place in the network, a node only needs to update the routing table with an alternative 'next forwarding node ID' towards a destination node.

The cycle of balancing interference impact using a maximum disturbed node threshold in MIR-VA is shown in Figure 63. In this cycle, whenever the immediate node density of a transmitting node rises, the DIIP of MIR will check the total number of this node's immediate neighbours, local disturbed nodes denote as DN_{Local}, and compare it with the maximum disturbed nodes, denote DN_{max}. It will then adjust this node's transmit power accordingly to maintain an optimal number of DN_{Local} within its LCG, and hence adjusting the degree of interference impact.

7.3 MIR-VA Performance Evaluation

This section shows the network performance based on the calculation of C_p and C_n using simulation results of a network model that was developed in OPNET. These results were then plotted using MATLAB in various formats.

7.3.1 Simulation model

The performance of the MIR-VA algorithm was investigated using a simulation test bed of 36 nodes distributed in a square topology approximately 60km^2, with a basic hop length of 3 km apart. The network model has three basic scenarios, i.e. square, chain, and hexagonal, which are similar to the network scenarios shown in Figure 48.

The network model has been simulated in four separate configurations with the local disturbed node thresholds (i.e. maximum disturbed node threshold) of n_d set to 5, 10, 15, and 20 respectively, and the Shortest-Path Routing (short for SP), MIR, SP with variable transmit power plus asymmetrical routing (SP-VA) and MIR-VA routing algorithms being simulated for comparison. Each node model, as illustrated in Figure 64, was configured to be capable of transmitting at a maximum power level, P_m, which is restricted by the n_d. This creates an A_i that is twice the size of A_t, due to each transmitting node communicating with its one hop neighbours, but interfering as far as two hops away.

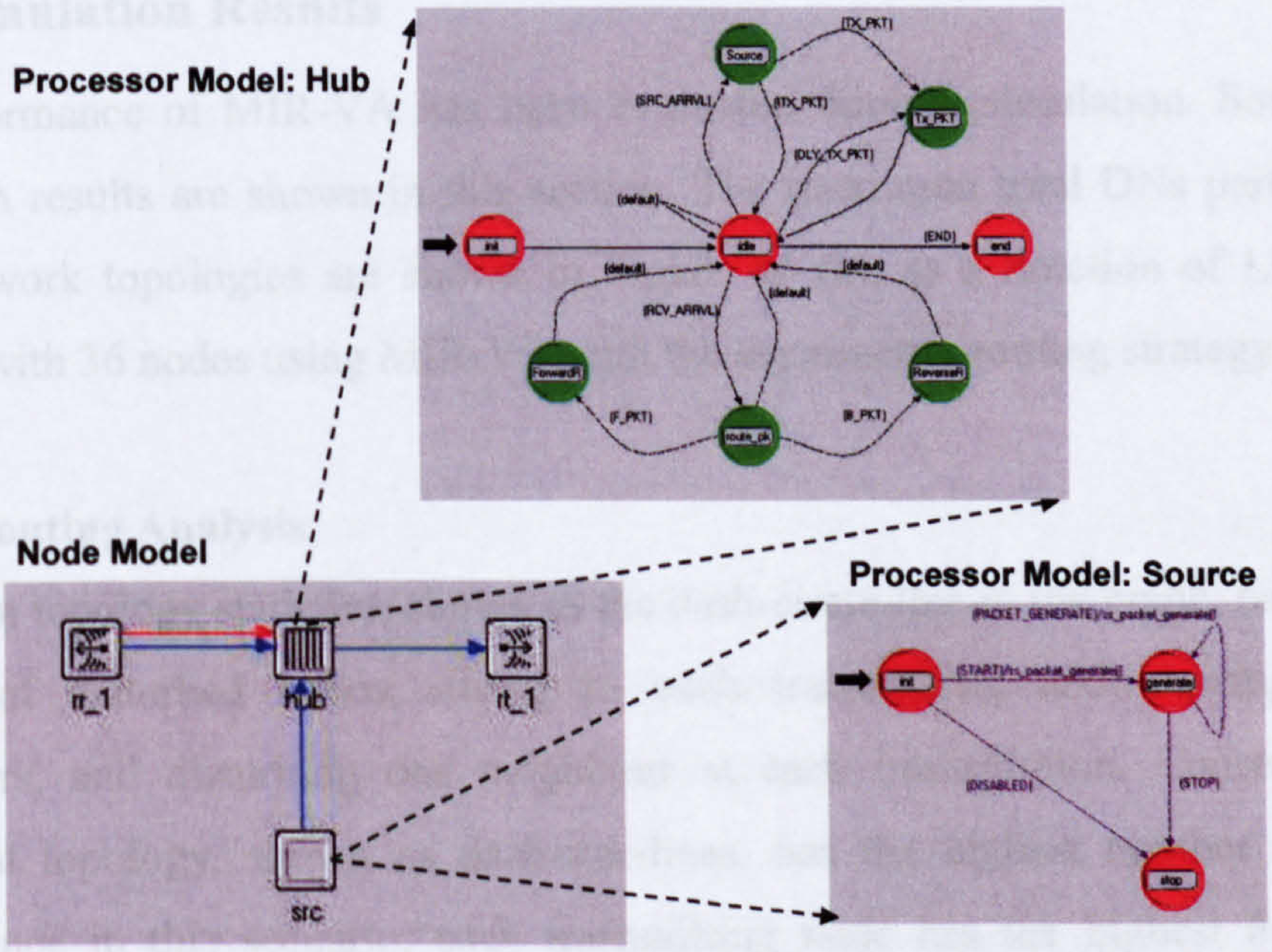


Figure 64 Illustration of Node Model used in OPNET simulator

Each node model using MIR-VA selects a route based on the route weight, R_W. F shows the functional diagram of a MIR-VA node model. When the SP algorithm is used, a route is selected based on the lowest number of accumulated disturbed nodes (DNs) as the routing criterion. The total DNs of each path between a pair of source and destination nodes were collected as a measure at the end of each simulation for analysis and comparison.

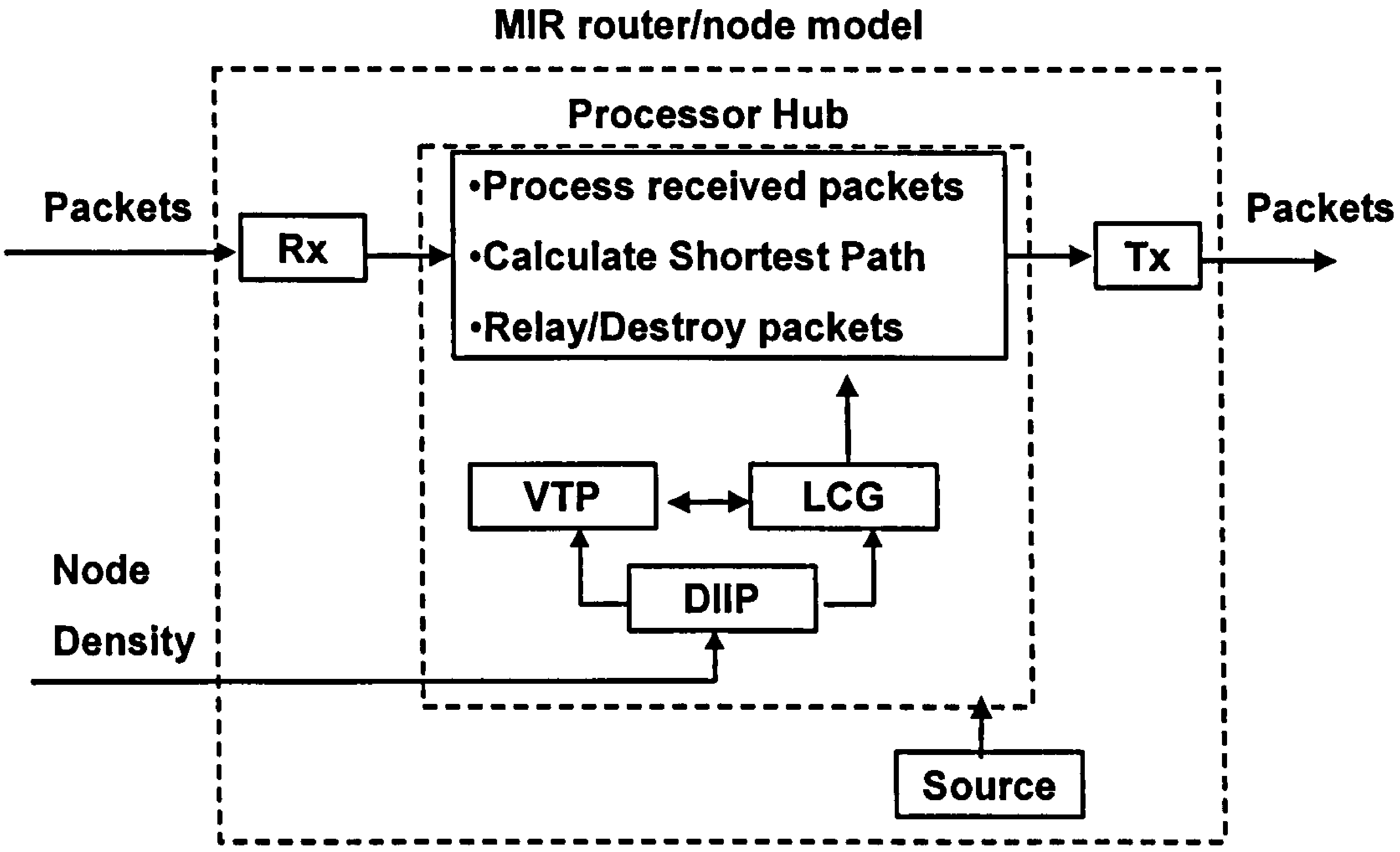


Figure 65 Logical Functions Diagram of a MIR-VA node.

7.3.2 Simulation Results

The performance of MIR-VA has been evaluated through simulation. Some selected simulation results are shown in this section. The maximum total DNs performance in three network topologies are shown in Figure 66 (a), as a function of Line-of-Sight distance with 36 nodes using MIR-VA with the asymmetric routing strategy.

7.3.2.1 Routing Analysis

The Chain topology statistics, shown as the dash-circle-line in the graph, has the lowest number of disturbed nodes, owing to each transmitting node having only two neighbours, and disturbing one neighbour at each transmission. Consequently, the hexagonal topology, shown as dash-star-lines, has the highest number of disturbed nodes, since in this scenario, each transmitting node has the highest 6 neighbours, whilst the square topology result shows the median performance is in the middle.

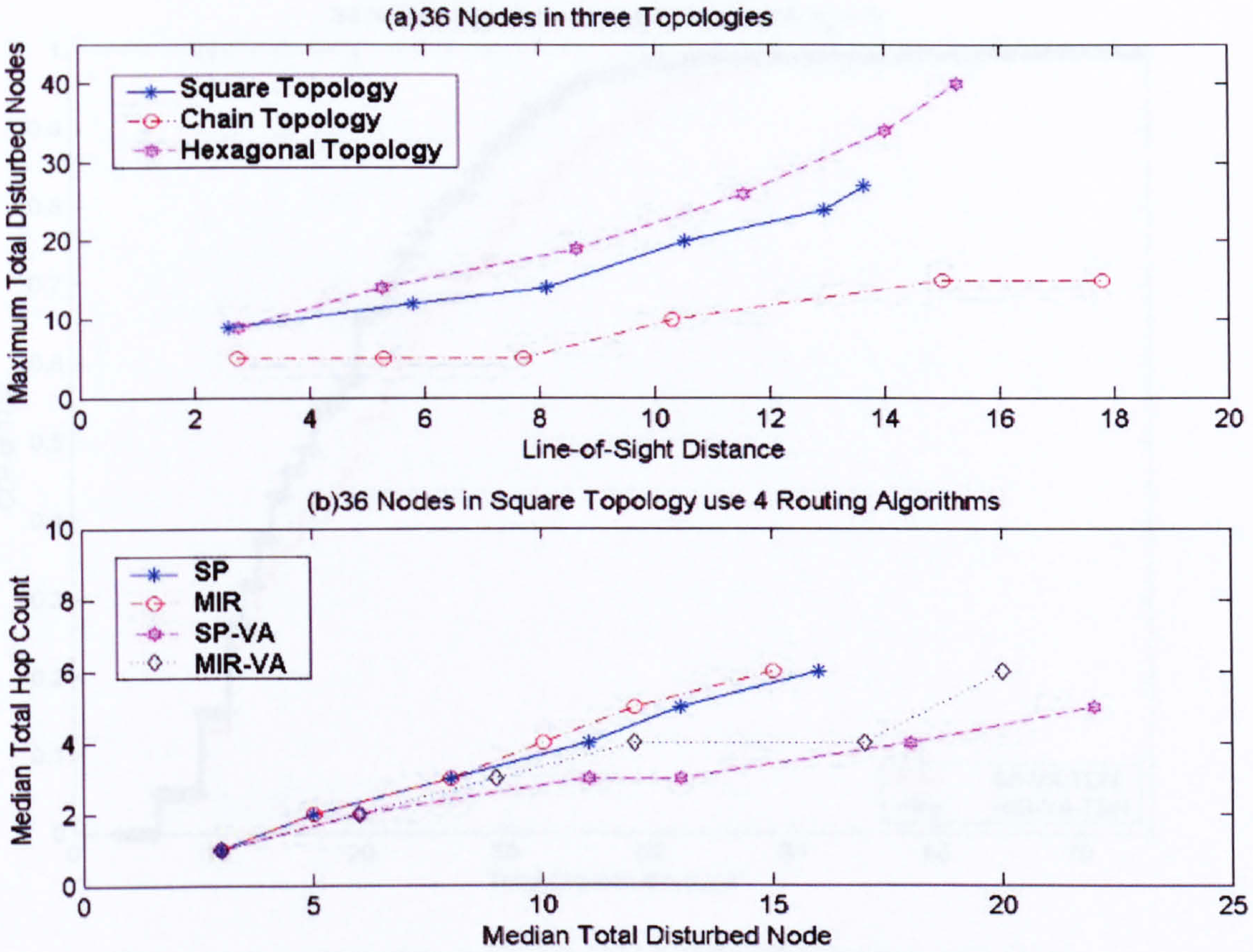


Figure 66 Disturbed Nodes and Delay Performance

(a) MIR-VA for 36 nodes for 3 topologies (b) 36 nodes in a square topology using 4 routing algorithms (SP, MIR, SP-VA, MIR-VA)

The performance of the square topology with 36 nodes is shown in Figure 66 (b), where the median total hop count is a function of median total DNs, and the model using the routing algorithms SP, MIR, SP-VA, and MIR-VA shows the variation matching the theoretical prediction.

For the same distance a packet has travelled, measured in hops, MIR has, on average, a smaller total number of DN than SP, and MIR VA has less DN than SP-VA with variable transmit power. For instance, at the average path length of 4 hops, the model using MIR results has an average of 10 disturbed nodes per path; for the model using SP, it shows 12 disturbed nodes on average per path; for MIR-VA, an average of 16 disturbed nodes per path is shown; and for SP-VA, an average of 18 nodes are disturbed per path. This verified the simulation with theoretical prediction.

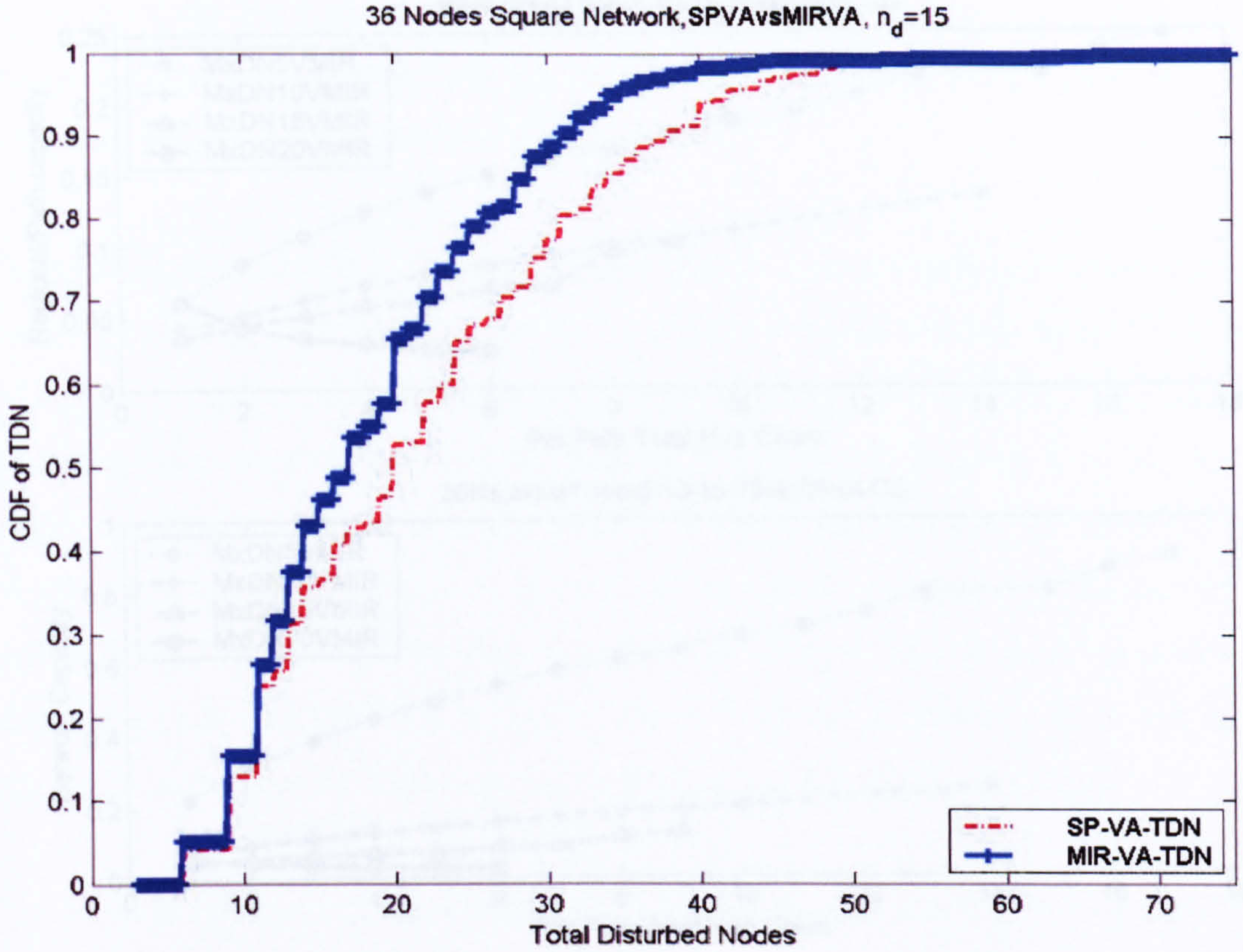


Figure 67 Cumulative Distribution Function (CDF) of Total Disturbed Nodes

MIR versus Shortest-Path (SP), Maximum n_d= 15.

Figure 67 shows the cumulative distribution function plot of TDN of SP- and MIR. The solid cross line on top is the CDF format of MIR's TDN for all routing tasks measured in the network. It shows, on the vertical axis, that 90% of the routing tasks using MIR disturbed fewer than 30 nodes per-path. In contrast, 90% of routing tasks using SP disturbed 37 nodes per-path, about 27% higher than MIR. This result also provides significant evidence showing that MIR-VA is capable of reducing higher ratios of DNs.

Figure 68 shows the path capacity, C_p, and network capacity, C_n, of the same square topology of the 36-node model, plotted as a cumulative distribution function against packet travelled distance measures in total Hop Count (HC) per path, which indicates the number of relays on the path.

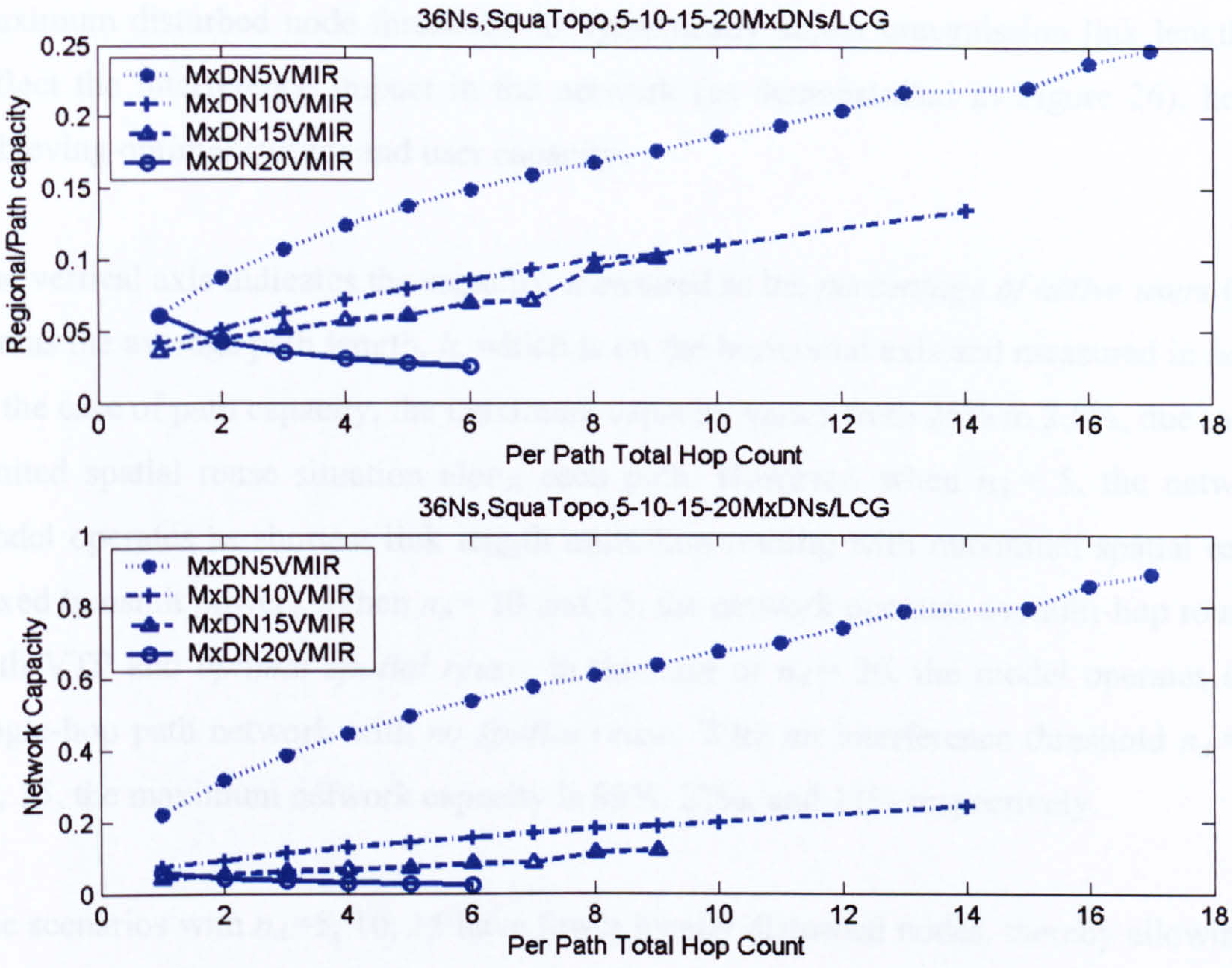


Figure 68 Network Capacity vs. Path Length for Maximum n_d=5, 10, 15, 20.

The MIR-VA scenario results in Figure 68 show that when the threshold of maximum disturbed node (DNmax) is set to a lower number, e.g. DNmax = 5 nodes, marked as MxDN5VMIR, measuring each transmission should disturb not more than 5 nodes. In this case the path and network capacity increases as the path length becomes longer, or in other words, the size of the network has increased. This is because as the number of nodes in the network is increased, more space becomes available to benefit or enhance the spatial reuse in the network. In this sense the network/system and user capacity will increase as the number of nodes increases exponentially, if the disturbed node threshold is chosen optimally.

This practical simulation result matches the mathematical prediction made by Toumpis and Goldsmith in [21] 2002, about the extended capacity boundary of WANET. In their investigation, the maximum system and user capacity was achieved by using a transmit power variation, in other words, it is the shortest link length routing. This can result in significant increase in multihop relaying traffic, delays, and routing overheads, which is not practical for implementation in WANET. Whereas MIR-VA used the balancing cycle of node density, local disturbed nodes, variable transmit power, and LCG maximum disturbed node threshold, to dynamically adjust transmission link length to reflect the interference impact in the network (as demonstrated in Figure 26), hence achieving optimal system and user capacity.

The vertical axis indicates the capacity, measured as the *percentage of active users* (%), versus the average path length, *h*, which is on the horizontal axis and measured in *hops*. In the case of path capacity, the maximum capacity varies from 25% to 2.5%, due to the limited spatial reuse situation along each path. However, when $n_d = 5$, the network model operates as shortest link length multi-hop routing with maximum spatial reuse (fixed transmit power). When $n_d = 10$ and 15, the network operates as multi-hop routing with VTP and *optimal spatial reuse*; in the case of $n_d = 20$, the model operates as a single-hop path network with *no spatial reuse*. With an interference threshold $n_d = 5$, 10, 15, the maximum network capacity is 89%, 22%, and 11% respectively.

The scenarios with n_d=5, 10, 15 have fewer locally disturbed nodes, thereby allowing a higher spatial reuse ratio, which results in higher and incremental network capacity. When $n_d = 20$, C_n decreases from 5% to 2%, since the local interference is high. As the number of relays increases, the network turns into a single-hop network with a low percentage of active users.

When the DNmax chosen is small, the transmit link length is not necessarily short, since the link length is controlled by the node density defined in LCG in conjunction with other influencing factors, such as local disturbed node density DN_{Local} sensed and measured use DIIP. This means the optimal link length is chosen by referring to the combination measure of interference impact, and balanced by the maximum disturbed node threshold DNmax, as illustrated in Figure 63, whereas Toumpis and Goldsmith's analysis proves only to depend on power variation and interference cancellation. There is no dynamic balancing mechanism in the mathematical analysis. This verifies and proves that the simulation result matches the theoretical predictions in section 7.2.5, about optimal interference impact measures providing increased path and network user capacity, whilst reducing multi-hop relay overheads.

On the contrary, if the disturbed node threshold DNmax is not chosen carefully or is too large, e.g. DNmax = 20, then both the path and network capacity will decreases as

Gupta and Kumar predicted in their paper [20], which indicates the system or channel capacity decreases at the ratio of $\frac{1}{\sqrt{N}}$ as the number of nodes increases.

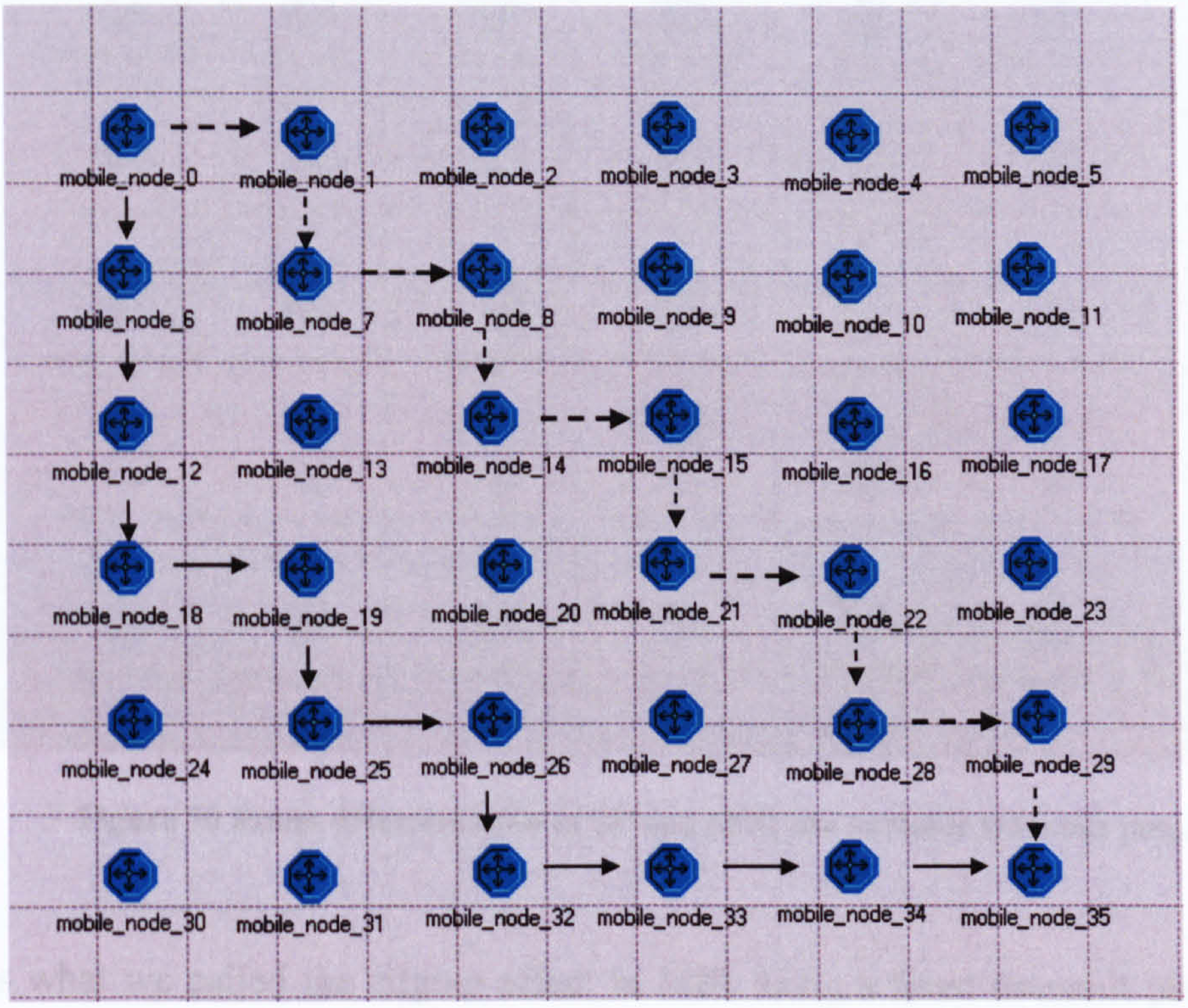


Figure 69 Route Determination of SP and MIR use fixed transmit power

Figure 69 demonstrates the route selection differences using shortest path routing (indicated by dash arrows lines) and MIR (indicated by solid arrows lines) with fixed transmit power that a packet will travel the same distance (10 hops), where SP disturbs 26 node along the path but MIR disturbs 20 nodes. SP uses the path travelled through the centre, which is shortest in distance, yet is easily congested. MIR tends to route towards the edge of the network, hence could divert traffic from the easily congested centre network.

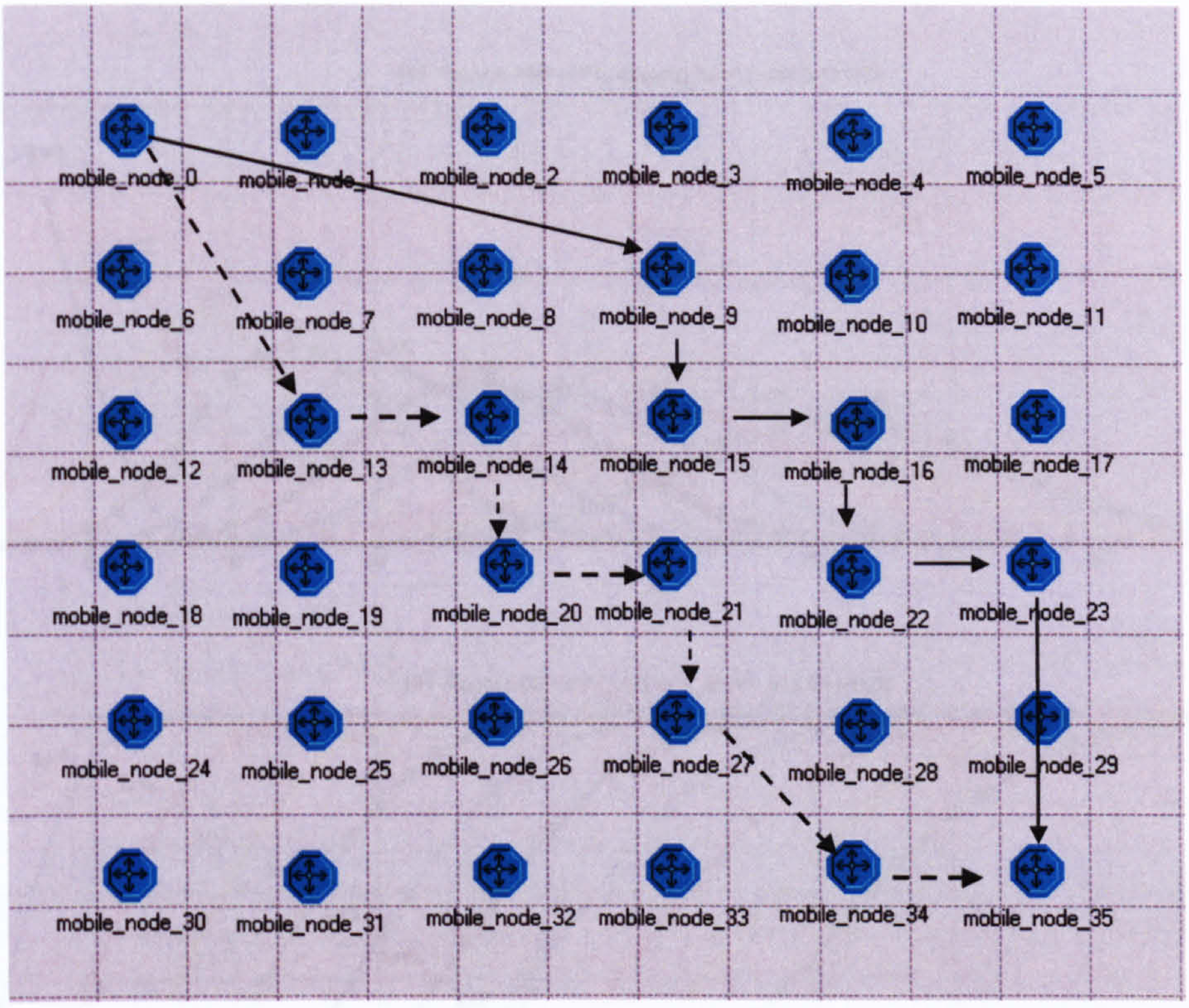


Figure 70 Route determination of SP and MIR use variable transmit power

This is what we called the edging effect in MIR using a fixed transmit power. This is resolved using a flexible DN_{max} threshold in conjunction with the use of variable transmit power in a distributed manner, so that nodes close to edge of the network will be able to access neighbours that are closer to the centre of the network, and preventing traffic flow aggregating around the edge of the network. Figure 70 shows the same network scenario with a different configuration, where the transmit power in each node is variable. The similar outcome show SP-VA disturbs 30 nodes, whilst MIR-VA disturbs 26 nodes.

7.3.2.2 Neighbour Node Distribution Analysis

In order to determine the optimal number of neighbours for a transmitting node, a random analysis has been carried out to investigate the probability of n_e neighbouring nodes occurring within a total of n_i interfered nodes (i.e. including the transmitting node n_t and a number of interfered neighbouring nodes n_e).

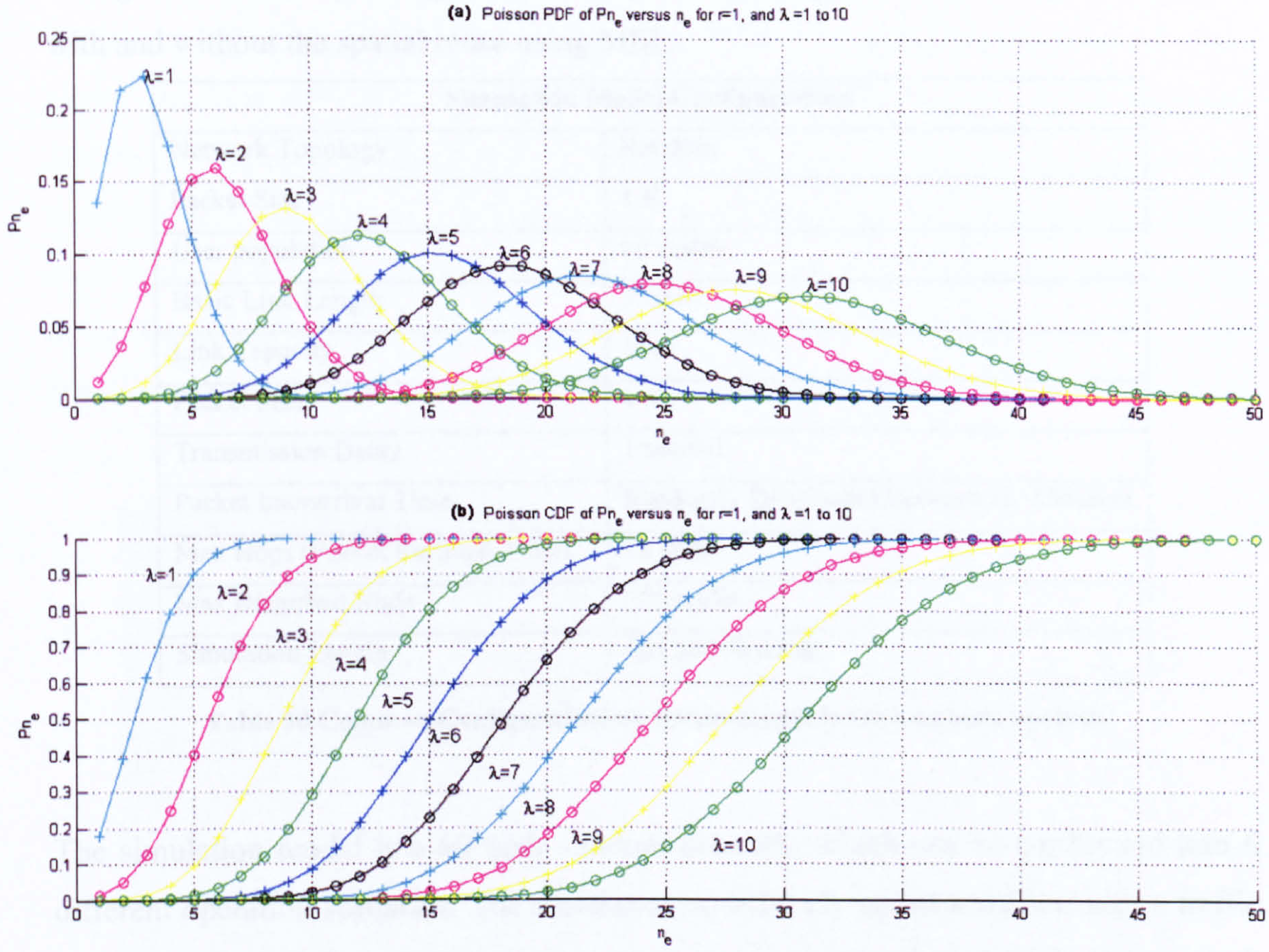


Figure 71 Poisson Distribution of Neighbouring Node Density

(a) Pdf of Probability of n_e neighbours in n_i interfered nodes

(b) CDF of Probability of n_e neighbours in n_i interfered nodes

Assume a total of N (e.g. 60) nodes randomly located in an area and only one node transmitting (n_t) at a time. In this case the probability of having n_e neighbouring nodes interfering within n_t's transmitting range is shown in Figure 71, in which (a) and (b) are the Pdf and CDF format of the Poisson distribution for n_e respectively.

7.3.2.3 System Structural Analysis

The previous study and analysis in chapter 5 indicates that different network structures have a different impact on system throughput. In order to validate this finding, a system performance analysis is constructed using a network model test bed, which has a common setting as shown in Table 10, for 6 different scenarios, each of which is configured differently to represent centralized or distributed, single-hop or multihop, with and without the spatial reuse using MIR.

Simulation Model Configurations	
Network Topology	Random
Packet Size	1 K
User population	60 nodes
Basic Link Length	3 Units
Link Capacity	1 K
Packet Size	1 K
Transmission Delay	1 second
Packet Interarrival Time	Randomly Distributed Between (0~500secs)
Max Hops (Packet Residual Time)	15 hops
Max Disturbed Node	>25 nodes
Simulation Length	604,800 Seconds

Table 10 Common Configuration of system model for throughput analysis

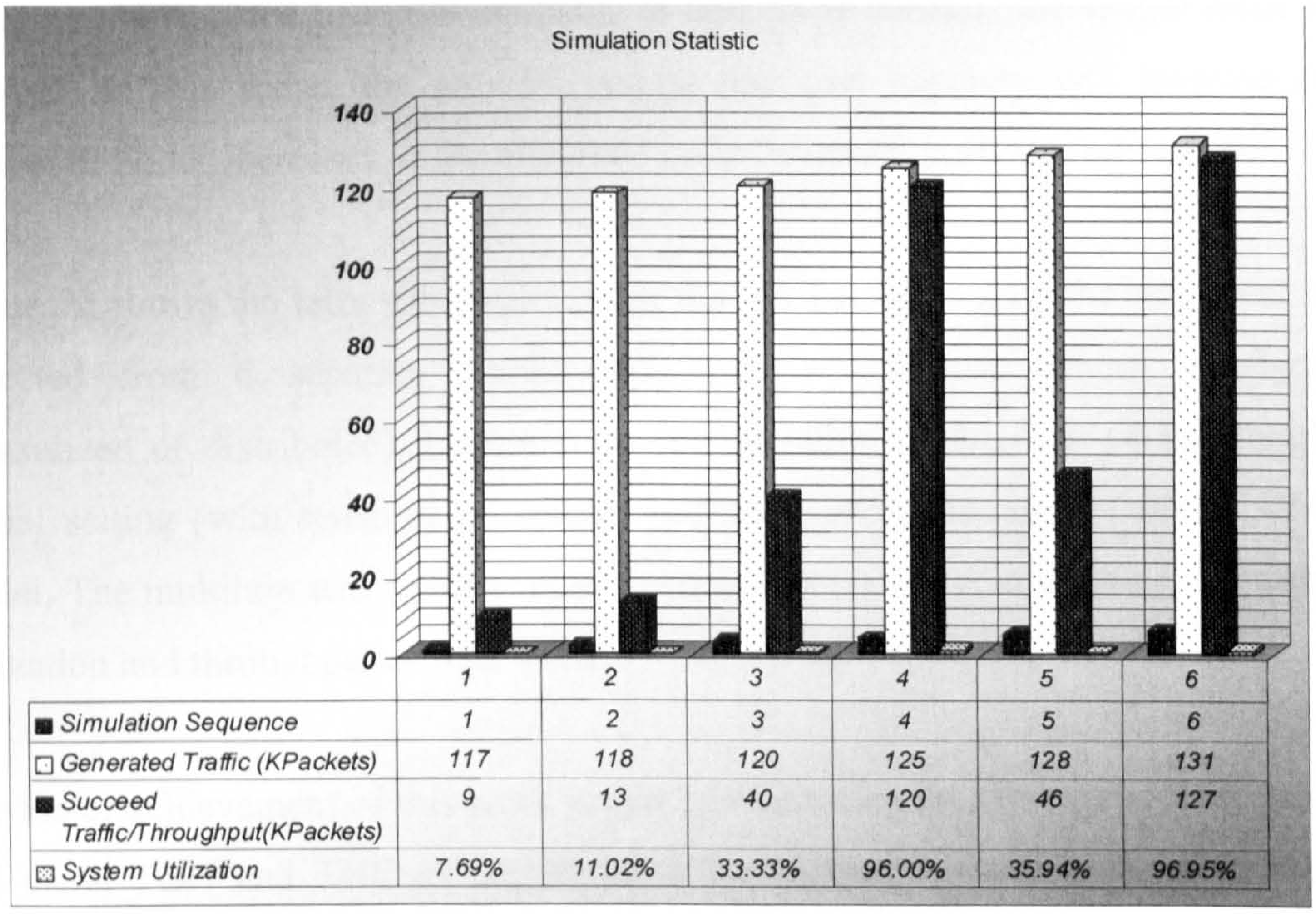
The simulation model is a 60 node random network, which can be configured into 6 different operation scenarios. The simulation specifically constructed includes a traffic counter to record the generation and successful delivery of packets in the network model, and the difference between these two measures will show the relaying or redundant traffic, and transmitted overhead in the network.

The simulation result is then collected to calculate system throughput and utilization for each system scenario. The statistical result is listed in Table 11, where it shows an expected result that matches and agrees with our earlier theoretical analysis in chapter 5 and the results presented earlier in this chapter.

Simulation Sequence	Representing System Configuration	Generated Traffic (K packets)	Successful Throughputs (K packets)	System Utilization
1	Centralized Single-Hop No Spatial Reuse	117	9	7.69 %
2	Centralized Single-Hop With Spatial Reuse	118	13	11.02 %
3	Centralized Multihop No Spatial Reuse	120	40	33.33 %
4	Centralized Multihop With Spatial Reuse	125	120	96 %
5	Distributed Single-Hop No Spatial Reuse	128	46	35.94 %
6	Distributed Multihop With Spatial Reuse	131	127	96.95 %

Table 11 Statistical results and simulation scenarios configurations

Among the 6 investigated system scenarios, 4 systems that transmit via single-hop, no spatial reuse, have in common a very low overall throughput that is less than 40%. This means less than 40% of the generated traffic (packets) successfully delivered to their destinations, in other words, the rest is overheads. In contrast, systems that transmit via multihop with spatial reuse, regardless of whether it is centralized or distributed control, total throughput reaches over 96%, which means the relaying burden (overheads) in these two scenarios has been reduced to minimum. Figure 72 demonstrates a more intuition comparison of these 6 system scenarios.



■ Simulation Sequence	1	2	3	4	5	6
□ Generated Traffic (KPackets)	117	118	120	125	128	131
■ Succeed Traffic/Throughput(KPackets)	9	13	40	120	46	127
▨ System Utilization	7.69%	11.02%	33.33%	96.00%	35.94%	96.95%

Figure 72 Statistical result for 6 Simulation with different network structures

7.4 Conclusion

This chapter has demonstrated the improved version of MIR with variable transmit power at each wireless node. MIR makes route select decisions depending on a minimum number of disturbed nodes accumulated along a path. When combined with a variable transmit power strategy, which sets the power level according to the number of nodes that are disturbed locally, a high network capacity can be achieved. The theoretical *network capacity* is derived based on the concept of *time sequenced* *interference regions* and the *path capacity*. The performances of the network in four scenarios are compared and show how network capacity can be improved.

The performance of MIR-VA was compared with simulations that use the same configuration, but different routing algorithms, i.e. shortest path (SP) routing algorithm, with a combination of fixed-identical and variable transmit power in each node. The MIR-VA scenario result in Figure 68 shows that shorter link length results in fewer number of disturbed nodes, and consequently contributes to the increase of overall system throughput and capacity. This is because as the number of nodes in the network increases, more space becomes available to benefit or enhance the spatial reuse in the network. In this sense, the network/system and user capacity will increase as the number of nodes increases, if the disturbed node threshold is chosen optimally.

Figure 72 shows the later simulation result that further backed up the statistical results collected from 6 separate simulations, with a different control configuration (centralized or distributed), transmission configuration (single-hop or multihop), and spatial setting (with reuse or no reuse), operating under the same OPNET WANET model. The multihop with spatial reuse featured simulation result shows a high system utilization and throughput of over 96%.

The major achievement of this work is that by measuring interference and relay impact, and using TDN and THC as the feedback for making a routing decision, networks operating MIR-VA can achieve a very high ratio of spatial reuse, and the variable transmit power dynamically adjusting the surrounding node density in each node's LCG eliminating the negative interference impact, whilst optimising the spatial reuse, and thereby achieving high network capacity.

Chapter 8 Future Work

Contents

8.1 Introduction

The techniques and issues associated with WANET scalability and routing strategies have been under substantial investigation and development. This research work developed so far has concentrated on two phases of development for the Minimum Impact Routing (MIR) algorithm, which has been designed with an objective of reducing the interference impact in a multi-hop wireless environment for WANET. Previous research on WANET routing used either highly abstractive assumptions (sometimes not even practical), or dealt with many practical problems. However, so far there is no single technique or algorithm that could provide a comprehensive solution for the scalability problem in WANET.

Pure MIR has established the foundation of this research using a Distributed Interference Impact Probing (DIIP) technique for measuring local neighbouring node density, which is strongly related to the interference impact measured within a one-hop distance of any transmitting node. This resulted in a measured number of local Disturbed Nodes (DN). These DNs then formed a Local Communication Group (LCG), in which the optimal number of DNs within each LCG is associated with a transmission link length. A series of simulations have been conducted in relation to finding out an optimal link length to form an optimal size LCG. However, the results of these simulations conclude that a dynamic environment such as WANET, requires an adaptive strategy to measure and adjust transmit power in order to optimise the tradeoffs, which is between the need to use high transmit power to achieve better connectivity and lower relays, and the demand of lower transmit power to reduce interference or conserve energy.

The MIR with Variable transmit power and Asymmetric routing (MIR-VA) incorporates variable transmit power with DIIP, LCG and asymmetrical routing. It thereby created an interference impact and connectivity balancing cycle as shown in section 7.2.6. In this mechanism, the balancing factor is a maximum disturbed node (DN_{max}) threshold, which directs the VTP to change the size of a LCG in selection to the measure result from DIIP. This architecture enables a transmitting node to alter the size of its LCG referred to as DN_{max}, hence controlling the interference impact region according to the local node density.

This research work examined WANET's capacity from a new perspective, in terms of interference impacts. However, the similarity between the proposed routing algorithms suggests the previous technique for connectivity was either too slow to converge, or too complicated to implement. A comprehensive routing strategy should be a simple routing algorithm that incorporates a mechanism that provides better scalability in WANET. The following sections discuss the possible techniques and schemes that MIR could or should incorporate, in order to enhance scalability in WANET whilst remaining simple to implement.

8.2 Physical Layer Improvement

8.2.1 Accurate Probing

The use of different measures of interference impact could be beneficial in future routing strategy development. The routing criteria, measures the associativity, interaction, and influence between nodes and could be investigated in further detail. For example, the SINR that gives a measure of signal strength can provide accurate detection on a physical layer. This can influence an LCG resizing during the operation. Other measures such as queue size and traffic load can also support the making of an LCG resizing decision. However, a single measure of these was proven to trigger routing oscillation [1, 8, 13]. A thorough study of issues associated with physical layer probing is helpful in order to understand and solve the scalability issue in relation to routing in WANET.

8.2.2 Autonomous Network Synchronization

In the case of WANET operating without the support of a fixed infrastructure, and the case where frequent time updates are not available, the synchronization between all the radio nodes could be a complex, yet important, issue that is essential for WANET implementation. Hence, a further study of synchronization related issues could be another direction for further development. Existing synchronization techniques are more or less designed for wireless terminals that are supported by a stationary central location, whereas the situation is a different for WANET. A novel biotechnology synchronization technique inspired by southern America firebug is an interesting extension of this issue [108].

8.3 MAC Layer Improvement

8.3.1 Perfecting Multiple Access Schemes

Effective multiple access schemes are essential for multi-user operation in multi-hop networks. A previous MIR-VA incorporates simple carrier sensing technique, Carrier Sensing Multiple Access / Busy Tone Sensing Multiple Access (CSMA/BTMA) in order to reduce collision, whilst maintaining simplicity to avoid overflow of routing overheads. Existing techniques proposed for WANET can be summarised as contention-based, e.g. random access, CSMA etc, in which collisions can be avoided, yet can not be eliminated; or alternatively passive access schemes, e.g. polling, in which case a controlling node polls through its neighbouring nodes and assigns access time slots for the appointed neighbouring node. The use of a practical yet stable enhancement should be incorporated. One solution could be to incorporate CSMA/CA with Request-To-Send and Clear-To-Send dialogue. The other could be to incorporate the polling scheme with MIR-VA.

8.3.2 Spread Spectrum Multiple Access Control

On the other hand, spread spectrum techniques are another alternative: the Frequency Sequence Spread Spectrum (FSSS) and Direct Sequence Spread Spectrum (DSSS) / Code Division Multiple Access (CDMA). These may simplify the frequency reuse in WANET, so that with the use of these frequency separation techniques, radio nodes in WANET can separate co-channel transmissions by physical distance or angles. The challenge in this issue is the demand for a high degree of inter-node cooperation. It would be worth looking at the potential benefits and drawbacks.

8.4 Network Layer Improvement

8.4.1 Highly Adaptive Routing Strategies

Current routing strategies generally have limited functionality, or are situation dependent. Therefore the development of a highly adaptive routing strategy is an essential trend as part of the routing evolution. Hence further investigation of the partition of networks and multi-point relay strategy (similar to the MPR used in OLSR), in which the whole network has only a proportion of active nodes to relay routing or data information, could be beneficial. In this case the total number of nodes needed to propagate and disseminate routing information could be significantly reduced.

MIR starts controlled broadcasting at the route discovery stage, and later uses unicast transmission inside each LCG, allowing a transmitting node to send towards a destination via a Last / Next Forwarding Node (LFN or NFN). This reduces packet duplication significantly when sending out a flooding packet. However it takes time to converge on an optimal return route. Therefore interactive operation of multi-path routing is a potential solution for this problem. Furthermore, the intermediate node using AODV can reply to a route request with full path length knowledge without forwarding the request packet all the way to its destination, providing it knows the rest of the path. This mechanism can reduce the existing routing relays of MIR by half, or more. Therefore an intelligent route discovery or relay mechanism should be able to improve the scalability of MIR, in terms of high efficient route discovery and relaying.

8.4.2 Coexistence of Co-channel users in Random Environments

Different systems use distinct standards, yet sharing the same spectrum can pose serious problems. Co-channel interference is manageable within the same system, however, such a type of interruption across a different system requires cross system cooperation. In some cases this could be difficult. Autonomous systems such as WANET have such an issue, e.g. coexistence of Bluetooth and WLAN. If WANET is operating in an urban area, the degree of noise and interference will be potentially higher among numerous wireless systems. Whether it can maintain communication, whilst not creating significant interference to others, is a complex and challenging issue. In such an environment, the network status is generally distributed in a random fashion. In this case, how WANET behave and performs in such an environment needs a deeper and a more thorough study. It is expected that an MIR based scheme could provide a solution, since it would be able to route around the system using the same channel. A more complex model should be developed.

8.5 System and higher Layer Improvement

8.5.1 Further investigation of Scalability in WANET issues

System level analysis of WANET is important for routing strategy development, owing to the fact that the burden of routing on WANET radio nodes is more complex than a stationary base-station. Scalability is strongly associated with WANET user capacity, apart from the channel throughput capability, and routing operations could be vitally influential, depending on how the networks maintain their connectivity. The complex interrelationship between VTP, node density, transmission link quality, spatial reuse etc, needs deeper investigation and understanding in practical user capacity for wide deployment of WANET. A version of variable power MIR that optimises the number of disturbed nodes within the LCG dynamically and autonomously should be investigated.

Furthermore, the system and user capacity derived and evaluated in this research was based on a unified network, with the assumption of a TDMA system. However, such thought can be extended to FDMA, and CDMA systems. Hence further analysis of system and user capacity in systems that use these multiple access schemes will be invaluable.

8.5.2 Traffic and Application oriented

The QoS issue is important to different application with distinct traffic types which need to be supported. WLAN provides two different types of MAC protocol, polling and CSMA/CD/CA, to support both contention based and contention free services. In some cases, the quality of the communication services is distinguished by the cost of service. Therefore, the best-effort, integrated services and differentiated service types can co-exist in the same system to satisfy different type of demands such as email and Internet services, voice and video on IP, live TV, and future broadband applications.

The implications of QoS constraints on MIR type schemes could be examined. One possibility would be to have several categories of disturbed nodes, where the level of a node could be set depending on the type of traffic transmitted by the node. This would provide a greater flexibility to divert route around highly vulnerable nodes.

Chapter 9 Conclusions

Contents

9.1 Summary and Conclusion

This thesis has investigated the techniques and issues associated with designing effective routing strategies to enhance user capacity in multi-hop Wireless Ad hoc NETworks (WANETs). WANET is well known for its potential to support a wide range of applications, however, its flexible and convenient nature also poses great challenges for implementation and deployment. Scalability of WANETs is a relatively new and unexplored research area for the international research community, and so far there has been no comprehensive single solution proposed for such a problem.

Routing is a network layer responsibility in conventional wireless systems, where these systems have sufficient resources such as bandwidth, power supply, and stationary central control locations etc. However, it is a completely different scenario in WANETs, where resources are precious and limited, and the lack of the support from a fixed infrastructure requires wireless terminals to carry the controlling burden of conventional central control units, such as routers, gateways etc. In such systems, communication quality is difficult to maintain due to shortened communications link length, which provides less signal strength in some cases; and terminals dominate the network management, hence demanding a higher degree of collaboration between wireless nodes and/or routers.

Chapter 2 reviews the characteristics and nature of WANETs. Routing in WANET faces the challenge of role conversion from centralized connectivity maintenance to a distributed interactive controller. The management and control burden shifts from a central location to a distributed terminal. This makes routing in WANET no longer just a technique for establishing and maintaining network connectivity, but more importantly participating in the system and user capacity enhancement, even management. Therefore WANET routing strategies not only relate to the network connectivity but are also associated with a very fundamental problem of WANETs, that of scalability.

In order to achieve this goal, a thorough background study of recent WANET research has been presented to highlight the challenges and issues associated with ad hoc networking, physical layer, MAC layer, network layer, system, and higher layers. This review of the literature identified some fundamental issues, such as interference, transmission power, contention control, and adaptive routing that need to be considered in relation to developing routing strategies for enhancing WANET system and user capacity.

Chapter 3 reviews WANET routing techniques, identifying influential issues in the particular routing environment, and studies the evolution of routing algorithms. The radio environment in WANETs can be dominated by *interference* due to the wireless and broadcast nature of those networks. In such an environment, the transmit power is the source of many other effects in the network. For instance, variation in transmit power will trigger variations in the radio link length, signal strength, signal-to-interference ratio, spatial reuse, number of routing relays and overheads, and power consumption. As a result of these changes, the network topology, connectivity, reliability, stability, scalability will be affected. Also important is the node density in the network. A higher density means higher connectivity but more interference. In contrast, a lower node density means less interference but weaker connectivity, therefore requiring higher transmit powers for longer radio links, which also results in higher interference. Therefore it is fundamentally important to balance the tradeoffs of finding the optimal link length that provides adequate signal strength for maintaining radio link connectivity, whilst producing the minimum amount of interference, and supporting a high degree of spatial reuse, thereby enhancing system and user capacity. This is the primary challenge that motivated this research.

In order to find a solid foundation for later routing strategy development, the evolution of routing algorithms is studied. Routing algorithms developed so far can be broadly summarized into two categories: centralized (static and dynamic) algorithms and distributed (proactive, reactive and hybrid) algorithms. The centralized schemes developed first with static routing mechanisms to support routing based on a stationary central location, e.g. routers, gateways etc., in which case the network topology is relatively static and the network status does not change often. Later on, dynamic routing mechanisms were developed to enable these central locations to accommodate network status changes, e.g. traffic load variation, link connection, and topology changes etc.

As communication systems develop, portable wireless terminals become capable of ever more functionality, and the developing concept about WANET gives wireless terminals the opportunity to operate with the function of a router. Distributed routing mechanisms were then developed to suit the demand of the decentralized type of routing in a proactive, reactive, or a hybrid manner. The proactive algorithms waste bandwidth to regularly update and maintain all routes leading to all destinations even when some are not in use, hence generating a large routing control overhead. Reactive algorithms, on the other hand, only initialise and maintain routes when needed. These algorithms take a long time to converge, however they cannot support low latency. Hierarchical algorithms are hybrid schemes that proactively maintain routes within a relatively small cluster of nodes, but route through the network in a reactive manner. These can be complicated to implement. The most important outcome from this study is that understanding of efficient routing strategies should have the feature of simplicity and adaptability. New routing algorithms were subsequently developed following this principle.

Chapter 4 looks at simulation techniques, analysis or validation methods associated with routing algorithm design and performance evaluation, together with discussions of advantages and disadvantages of these methods. The simulation results represent primarily raw statistical data collected from a simulation model constructed in the industry standard communication simulator, OPNET. This data is then processed in MATLAB for producing graphical evidence, used later for performance evaluation. Routing algorithm modelling in a wireless environment involves Proto-C source code (a programming language similar to C) programming in a hierarchical simulation model. This simulation model can be investigated and modified from the top network level down into the node level. Below this is the process level where the functional components of a node are then described, and finally in each processor there is a state-transaction-diagram, in which each state contains fundamental source code. This wireless simulation model required modification to the radio transceiver pipeline in OPNET, in order to tailor the simulation environment to suit the routing algorithm modelling demands.

Chapter 5 examines and develops the foundation of a routing algorithm simulation model as a test bed for later design and improvement. A new routing strategy, Minimum Impact Routing (MIR), is introduced. Two fundamental issues, Distributed Interference Impact Probing (DIIP) and distributed Media Access Control (MAC), associated with later MIR design are introduced in this chapter. This basic routing simulation model uses the distance a packet has travelled (measured in hops) as the routing criterion and later for calculation of the shortest path. A cable-based and a wireless network model were developed for comparison. The construction of this model follows the principle of providing a fundamental routing model with high simplicity.

Chapter 6 then describes how MIR was developed to minimize interference and enhance system and user capacity in WANET. It is not possible to eliminate interference to a defined level, since interference will always exist, however it can be reduced to an acceptable level for successful communication. It is shown how MIR combines essential factors, such as interference impact probing (using DIIP), distributed contention and topology control (use MAC and LCG) in order to reduce the impact of interference on system and user capacity, and reduce routing relay overheads. MIR focuses on the consequences of interference rather then the interference itself, and uses DIIP to measure such impact around any transmit nodes, thus quantifying the interference impact with the measure of Disturbed Nodes (DN), which reflects the interference impact on local immediate neighbours (within the reach of one hop).

Pure MIR established the foundation of this research by building on the DIIP mechanism for measuring local neighbouring node density, which is associated with the interference impact measurable within a one-hop distance of any transmitting node. The result of probing measures such as the number of local DNs, which in this case is used for defining the size of a Local Communication Group (LCG). The optimal number of DNs within each LCG is associated with transmission link length. A series of simulations have been conducted to find an optimal link length to form an optimal LCG. However, the results of these simulations conclude that in a dynamic environment such as a WANET, an adaptive routing strategy is required to measure and adjust transmit power dynamically, in order to better compromise the tradeoffs between the need to use high transmit power to achieve better connectivity (or lower relays) and the demand of lower transmit power to eliminate interference or conserve energy.

Chapter 7 introduces further developments on the pure MIR scheme, specifically incorporating Variable Transmit Power (VTP) and an asymmetrical routing mechanism into the MIR framework. MIR-VA is the improved version of MIR and dynamically controls interference impact by autonomously adjusting each node's transmit power level, based on the comparison result of a local DN measured by DIIP with a pre-determined maximum DN per LCG. The DNmax acts like a threshold, to balance the variation of interference impact (yielded by neighbouring nodes) and the variation on link length, by selecting an appropriate transmit power level. In this case, each wireless node can dynamically control the surrounding neighbour node density and hence control the interference impact among them. The performance evaluation of MIR-VA verified that with an optimally chosen maximum disturbed node threshold (DNmax), a network capacity of 90% of nodes could actively make concurrent transmissions over a period of time.

Inspired by spatial reuse theory analysis conducted originally for MIR-VA, the system and user capacity of a unified network (measured as the percentage of active users) was derived based on the concept of Time Sequenced Interference Regions (TSIR) and the *path capacity*. The performance of MIR-VA was compared with simulation models that use the same configuration but different routing algorithms, i.e. Shortest Path (SP) routing algorithm with a combination of fixed-identical or variable transmit power in each node. A number of comparisons were conducted in four routing scenarios, and it is shown how the network capacity can be improved by using a lower value of DNmax with variable transmit power. This results in the highest path and network capacity, measured by the number of active transmitting nodes. The variable transmit power dynamically adjusts the surrounding node density of each node to reduce the negative interference impact, whilst optimising spatial reuse, thereby achieving higher network capacity. Both path and network capacity are proportional to a spatial reuse ratio calculated using the spatial reuse theory and TSIR conceptual model. This measure reflects the effect of spatial reuse and packet relays in the WANET simulation mode.

When the DNmax chosen is small, the transmit link length is not necessarily short, since the link length is controlled by the node density defined in LCG in conjunction with other influencing factors, such as local disturbed node density DN_{Local} measured use DIIP. This means the optimal link length is chosen by referring to the combination of interference impact and maximum disturbed node threshold DNmax, as illustrated in Figure 63. This mechanism agrees with the mathematical analysis made by Toumpis and Goldsmith suggesting transmit power variation and interference cancellation can improve system and user capacity asymptotically. However, differently from their model, MIR-VA uses a dynamic interference impact balancing mechanism to achieve the same goal. This is verified and proves that the simulation result matches the theoretical predictions in section 7.2.5, with regard to the optimal interference impact measures providing increased path and network user capacity, whilst reducing the multi-hop relay overheads.

On the contrary, if the disturbed node threshold DNmax is not chosen carefully (either too small or too large, e.g. DNmax = 20), then both the path and network capacity will decrease asymptotically. This agrees with what Gupta and Kumar predicted in their paper [20], indicating the system or channel capacity decreases at the ratio of $\frac{1}{\sqrt{N}}$ as the number of nodes increases asymptotically. The simulation results have demonstrated that the interference impact balancing cycle could achieve improved system and user capacity, based on the measure and calculation of node interactions in a spatial reuse theory in section 7.2.3. However, the channel capacity was set to a logical assumption of unified traffic with optimal throughput of 1 Erlang for the simulation model. Hence a detailed study on channel modelling in conjunction with the achieved MIR-VA model is predicted to be invaluable for further investigation in order to make it closer to the real wireless network environment implementation.

This research work has only developed the foundations of a new adaptive WANET routing strategy for enhancing system and user capacity, and further extension of this work is essential. Chapter 8 shows the future development details of MIR-VA, and improvements that can be made to the physical, MAC, network, and system level aspects.

9.2 Original contributions

The novel and original contributions of this work presented in this thesis can be summarized into three distinctive areas: MIR related; scalability related; and analytical or theoretical issues

9.2.1 Minimum Impact Routing (MIR) Related

- An original approach to implementing adaptive interference-aware WANET routing has been invented (chapter 6). This technique quantifies the impact of co-channel interference, and uses the concept of disturbed nodes (DN) as a measure of the impact on making routing decisions. In this way, routing algorithms can enhance the spatial reuse and scalability of the system, whilst reducing the negative impact of co-channel interference on system and user capacity.

- A novel distributed interference impact probing (DIIP) mechanism has been developed to measure the DNs of each transmitting node. This measure of DN defines the size of a LCG for each node. In this way, DN can define the neighbouring node density of this node, and the size of its interference impact region.

- A new family of WANET routing protocols have been developed to incorporate the novel DIIP technique (chapter 5 and 6), with VTP, LCG, and asymmetrical routing (chapter 7). The MIR algorithms (pure MIR and MIR-VA) are able to make routing decisions dynamically based on the measure of interference impact, thereby improving spatial reuse and scalability of the system, and furthermore enhancing system and user capacity.

- A new routing protocol has been introduced (chapter 6). Pure MIR uses the novel accumulated interference impact measured along a path (Total DN per path) as the routing criterion for making routing decisions. This reflects the impact along any path, and also allows traffic to be directed away from highly congested areas, such as the centre of a network, where all shortest path type routing algorithms tend to direct traffic. MIR chooses the path with the minimum interference impact. In addition, networks using MIR tend to disturb neighbouring nodes that are not communicating with the transmitting node as little as possible, therefore maximizing the spatial reuse in the system, allowing more concurrent transmissions to take place. This work resulted in a publication for an international conference, WPMC'04.

- A new routing strategy combining VTP, LCG, DNmax, and asymmetrical routing mechanisms has been developed (chapter 7). MIR-VA associates a maximum DN threshold (DNmax) with VTP, DIIP, and LCG, to dynamically adjust the size of the LCG of each node, using power variation, based on the result of DIIP measured local DN (DN_{Local}). In this case, the transmission link length (defining the size of a LCG) changes dynamically depending on the local node density (DN_{Local}) measured using DIIP. This results in asymmetric links along a path with a reduced number of relays. Therefore, additional to the pure MIR's beneficial nature, MIR-VA can reduce the number of relays along the path, and therefore reduce both the relaying delays and relay overheads. This work resulted in a publication for an international conference, WiCOM'06.

- A simulation test bed has been developed in OPNET, dedicated to the MIR routing algorithm family, with a series of configurations and scenarios (chapter 4,5,6,7).

9.2.2 Scalability Related

- The major achievement of this work is that of measuring interference and relaying impact to dynamically adjust link length to alter the impact range, and using DNs for making routing decisions. As a result, networks using MIR and later MIR-VA can achieve a higher spatial reuse ratio. This serves to enhance system and user capacity, and improves scalability in WANET.

- A novel derivation of spatial reuse, path capacity and network capacity has been carried out. Later analysis of spatial reuse theory and the TSIR analytical model (chapter 7) associated the interference impact measuring mechanism with the system and user capacity analysis.

9.2.3 Analytical and Theoretical issues

- A novel interference impact analysis and measurement concept has been developed (chapter 6,7). The distinctive DIIP technique quantifies interference impact in a WANET environment.

- A unique spatial reuse analytical model has been developed in order to understand the interaction between concurrent transmitting nodes in a unified wireless environment (chapter 7).

- A distinctive Time Sequenced Interference Region (TSIR) model has been developed to analyse the concurrent transmitting node behaviour over a period of time (chapter 7). A novel path capacity analytical model has been derived based on the analysis of the TSIR model. Consequently, a distinctive network capacity analytical model has been developed based on the path capacity model (chapter 7). A detailed performance evaluation through simulation and analysis of the MIR routing algorithms with DIIP measures has been carried out in chapter 6 and 7. Simulation results are analysed in MATLAB.

Appendix I. Preliminary Knowledge

A.I.1 Probability Density Function

The probability density function is an integral of the probability mass measured over a given interval. The probability mass associated with an interval can also be obtained by computing the difference in the CDF for the upper and lower limits of the interval. As interval widths become infinitesimally small, it can be seen that the PDF is the derivative of the CDF with respect to the outcome (i.e., ordinate) variable.

The relationship between a PDF and a CDF is in fact the basis for the method used by the Analysis Tool to compute PDFs. A CDF is first computed, and a differentiation is performed to construct a PDF. The original statistic data is discrete, so that differentiation is performed in an approximate manner. This means dividing probability mass associated with an interval by the interval's width. The difference between two consecutive CDF values is divided by the difference in the corresponding ordinates. The resulting value is taken as the density associated with the interval and is placed at the intervals lower limit. Thus, if a statistic contains two consecutive ordinate values y1, and y2, the PDF will be computed as follows:

$$PDF(y1) = \frac{CDF(y2) - CDF(y1)}{y2 - y1}$$

An immediate consequence of this computation method is that PDFs can have extremely large values when the input statistic has distinct but closely spaced ordinate values. This is because the $(y_2 - y_1)$ difference becomes small.

If input statistic ordinate values are unevenly spaced (i.e., some very small differences exist, but also some significantly larger ones), PDFs can have a discontinuous appearance with certain density values dwarfing others. In such cases, PDFs is not as useful as the PMF or histogram operations.

A second consequence of this calculation is that the PDF contains one less entry than the CDF due to the fact that no forward-looking difference can be calculated for the final (i.e., maximum) ordinate value.

Finally, the integral of the PDF statistic, which can be computed using the appropriate filter, produces a statistic, which is identical to the CDF in its shape. However, the initial value of the CDF is lost in computing the PDF, which means the two statistics differ by a constant. This difference is particularly noticeable when the original statistic has a small number of distinct ordinate values, since the CDFs value for the minimum ordinate is at least the reciprocal of this number (i.e., this is the probability mass associated with the first ordinate value). Total area under the curves of a PDF graph is always 1, which indicates 100% when expressed as a percentage.

A.I.2 Cumulative Distribution Function

Like the probability mass function, the cumulative distribution function (CDF) of a statistic is related to the likelihood of occurrence of the statistic's ordinate values. However, rather than provide the probability mass of each ordinate's occurrence, the CDF shows the accumulated probability mass of all ordinates less than or equal to a particular ordinate. This form of presentation is useful when particular ordinate value thresholds are of interest, such as when determining the likelihood of receiving a message whose delay exceeds a particular value. For example, the probability of receiving a packet with a delay of 20 ms must be no greater than 0.1.

The computation of a CDF resembles that of a PMF in the sense that proportions for each ordinate value in the original statistic are computed. The same weight, which is the reciprocal of the number of entries, is attributed to each entry. Thus, if there are 100 entries, each entry has a weight of 0.01, and if there are five entries whose ordinate value is y, the ordinate y has a total probability mass of 0.05. The entries of the CDF are constructed by positioning the distinct ordinate values of the original statistic in increasing order on the abscissa with one entry for each such value. The CDF value for the initial entry is simply the probability mass of the corresponding ordinate. The CDF value for the second entry is equal to the CDF value of the first entry augmented by the probability mass of its corresponding ordinate value and so on. The CDF is essentially a running sum of the values of the PMF. Two simple properties of the CDF result from the method of computation described as follows:

(1) Since each CDF value is computed by adding a positive probability mass to the previous value, CDFs are monotonically increasing.

(2) Since the sum of all probability masses must add up to unity, all CDFs must have a final value of 1.0. This also makes sense under the definition of the CDF because one would expect the likelihood of obtaining an ordinate value less than or equal to the maximum ordinate value to simply be 1.0.

A.I.3 Carrier Sense Multiple Access

Carrier Sense Multiple Access (CSMA) is a MAC protocol that a node verifies the absence of other traffic before transmitting on a shared physical medium, such as a cable bus, or a band of spectrum. "**Carrier Sense**" describes the fact that a transmitter listens for a carrier wave before trying to send. That is, it tries to detect the presence of an encoded signal from another station before attempting to transmit. If a carrier is sensed, the node waits for the transmission in progress to finish before initiating its own transmission. "**Multiple Access**" describes the fact that multiple nodes send and receive on a medium. All other nodes using the medium generally receive transmissions by one node.

Concurrent transmission by multiple nodes results in frame collisions. The multiple transmissions interfere with one another so that all are garbled and receivers are unable to distinguish the overlapping received signals from each other. It is impossible to entirely prevent collisions in CSMA networks, but there are three ways to address them:

In pure CSMA, only the carrier sense is used to avoid collisions. If two nodes try to send a frame at nearly the same time, neither detects a carrier so both begin transmitting at the same time. The transmitters do not detect collisions, so transmit the entire frame (thus wasting the bandwidth used). Receivers cannot distinguish between collisions and other sources of frame errors, so collision recovery relies on the ability of the communicating nodes to detect frame errors and invoke an error recovery procedure. For example, the receiver may not send a required ACK, causing transmitters to time out and retry.

In Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), each node must inform other nodes of an intent to transmit. When the other nodes have been notified, the information is transmitted. This arrangement prevents collisions because all nodes are aware of a transmission before it occurs. However, collisions are still possible and are not detected so have the same consequences as in pure CSMA.

In Carrier Sense Multiple Access with Collision Detection (CSMA/CD), sending nodes are able to detect when a collision occurs and stop transmitting immediately, backing off for a random amount of time before trying again. This results in much more efficient use of the media since the bandwidth of transmitting the entire frame is not wasted. However, it is not possible with all media (e.g., radio), and requires extra electronics (not really an issue with today's technology, but one reason Apple used CSMA/CA-based Local-Talk instead of the then much more expensive Ethernet).

In Carrier Sense Multiple Access with Bitwise Arbitration (CSMA/BA), all of the nodes on the bus are assigned an identification number or priority code. When a collision occurs, one of the nodes that are attempting to send at the same time will be given priority to transmit according to its identification number or priority code (as opposed to waiting a random amount of time and then retransmitting, as in CSMA/CD).

Erlang In communications, a unit of telecommunication traffic intensity determined by the product of the number of calls, carried by the circuit in one hour, and the average duration of the call in hours. Traffic Intensity (Erlang) = Number of calls (carried by circuit in 1 hour) x Average Duration per call (measured over a period of time).

Near-Far problem is the problem of very strong undesired users' signals overwhelming the weaker signal of the desired user, can severely decrease performance.

A.I.4 Queueing Systems Brief Summary

Detail of the table refer to [101]

	Type of Queue / Performance Measures	M/M/1	M/M/∞	M/M/m (Erlang C)	M/M/1/K	M/M/m/m (Erlang B)	M/M/1//m	M/M/∞//m	M/M/m/K/M Engset Distribution
age nterarrival ime $t=\frac{1}{\lambda}$	Arrival Process (Offered flow or load)	$\lambda_k=\lambda \;\; k=0,1,2...$	$\lambda_k=\lambda \;\; k=0,1,2...$	$\lambda_k=\lambda \;\; k=0,1,2...$	$\lambda=\begin{cases}\lambda & k<K\\ 0 & k\ge K\end{cases}$	$\lambda_k=\begin{cases}\lambda & k<m\\ 0 & k\ge m\end{cases}$	$\lambda_k=\begin{cases}\lambda(M-k) & 0\le k\le M\\ 0 & otherwise\end{cases}$	$\lambda_k=\begin{cases}\lambda(M-k) & 0\le k\le M\\ 0 & otherwise\end{cases}$	$\lambda_k=\begin{cases}\lambda(M-k) & 0\le k\le K-1\\ 0 & otherwise\end{cases}$
verage ervice ime $x=\frac{1}{\mu}$	Service Process (Link Throughput)	$\mu_k=\mu \;\; k=0,1,2...$	$\mu_k=k\mu \;\; k=1,2,3...$	$\mu_k=\begin{cases}k\mu & 0\le k<m\\ k\mu & k\ge m\end{cases}$	$\mu_k=\mu \;\; k=0,1,2...K$	$\mu_k=k\mu \;\; k=1,2,...,m$	$\mu_k=\mu \;\; k=1,2,...$	$\mu_k=k\mu \;\; k=1,2,...$	$\mu_k=\begin{cases}k\mu & 0\le k\le m\\ m\mu & k\ge m\end{cases}$
	Num. Of Server	1 Server	∞ Servers	m Server	1 Server	m Server	1 Server	∞ Servers	m Server
	Storage	∞ Storage	∞ Storage	∞ Storage	K Finite Storage	m Storage	∞ Storage	∞ Storage	K Storage
ystem Performance Measures	User/Customer Population	∞ Population	∞ Population	∞ Population	∞ Population	∞ Population	m Population	m Population	M Population
	Average/System Throughput:	$\overline{N}=\frac{\rho}{1-\rho}$	$\overline{N}=\frac{\lambda}{\mu}$	$\overline{N}=\frac{\lambda}{\mu}$					
	Delay	$T=\frac{\overline{N}}{\lambda}=\frac{\frac{1}{\mu}}{1-\rho}=\frac{1}{\mu-\lambda}$	$T=\frac{1}{\mu}$	$T=\frac{1}{\mu}$					
	Utilisation	$\rho=\lambda*\overline{\chi}=\frac{\lambda}{\mu}$	$\rho=\frac{\lambda}{\mu}$	$\rho=\frac{\lambda}{\mu}$					
	Probability Of "k" Success (operations; transmits;)	$p_k=(1-\rho)\rho^k$	$p_k=p_0\frac{\left(\frac{\lambda}{\mu}\right)^k e^{-\frac{\lambda}{\mu}}}{k!}$ $k=0,1,2,...$	$p_k=\begin{cases}p_0\frac{(m\rho)^k}{k!} & k\le m\\ p_0\frac{(\rho)^k m^m}{m!} & k\ge m\end{cases}$	$p_k=\begin{cases}\frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{K+1}}\left(\frac{\lambda}{\mu}\right)^{K+1} & 0\le k\le K\\ 0 & otherwise\end{cases}$	$p_k=\begin{cases}p_0\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!} & k\le m\\ 0 & k>m\end{cases}$ $p_m=\frac{\frac{\left(\frac{\lambda}{\mu}\right)^m}{m!}}{\sum_{k=0}^{m}\frac{\left(\frac{\lambda}{\mu}\right)^k}{k!}}$	$p_k=\begin{cases}P_0\left(\frac{\lambda}{\mu}\right)^k\frac{M!}{(M-K)!} & 0\le k\le M\\ 0 & K>M\end{cases}$	$p_k=\begin{cases}\frac{\left(\frac{\lambda}{\mu}\right)^k\binom{M}{K}}{\left(1+\frac{\lambda}{\mu}\right)^m} & 0\le k\le M\\ 0 & otherwise\end{cases}$	$p_k=\frac{\binom{M}{K}\left(\frac{\lambda}{\mu}\right)^k}{\sum_{i=0}^{m}\binom{M}{i}\left(\frac{\lambda}{\mu}\right)^i}$ $k=0,1,...,m$

Appendix II. Glossary and Abbreviations

A

ABR	Associativity Based Routing
ACK	Acknowledge Character
ADSL	Asymmetric Digital Subscriber Line
AL	Asymmetric Links
AODV	Ad hoc On-demand Distance Vector
ARPANET	Advanced Research Project Agency Network
ATM	Asynchronous Transfer Mode

B

BGP	Border Gateway Protocol
BP	Beacon Packet
BTMA	Busy Tone Multiple Access

C

CCITT	Consultative Committee for International Telephone & Telegraph
CGSR	Clusterhead Gateway Switch Routing
CSMA-CD	Carrier Sense Multiple Access with a Collision Detection
CSMA-CA	Carrier Sense Multiple Access with a Collision Avoidance
CTS	Clear To Send
CPE	Central Processing Equipment
CMIP	Common Management Interface/Information Protocol

D

DAG	Directed Acyclic Graph
DBF	Distributed Bellman-Ford
DCF	Distributed Coordination Function
DIIP	Distributed Interference Impact Probing
DN	Disturbed Nodes
DoD	Department of Defence
DPC	Distributed Power Control

DSDV	Destination Sequenced Distance Vector
DSR	Dynamic Source Routing
DSSS	Direct Sequence Spread Spectrum
DTC	Dynamic Topology Control ()
DVR	Distance Vector Routing
DVL	Distance Vector Lists

F

FBR	Flow-Based Routing
FHSS	Frequency Hopping Spread Spectrum
FP	Forward Path
FDD	Frequency Division Duplexing
FDM	Frequency Division Multiplexing
FDMA	Frequency Division Multiple Access
FRN	First Relay Node
FSM	Finite State Machine
FSSS	Frequency Sequence Spread Spectrum
FTP	File Transfer Protocol

G

GEO	Geostationary Earth Orbit
GRE	Generic Routing Encapsulation
GloMo	Global Mobile Information Systems

H

HC	Hop Count

I

IERP	Intra-zone Routing Protocol
IR	Interfere Range
IRR	Interference Range Ratio
IRD	Integrated Receiver Decoders
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector
IS-IS	Intermediate System-Intermediate System

IP Sec	Secure Internet Protocol (IETF)
IPv6	Internet Protocol version 6
ISDN	Integrated Services Digital Network

K

KPs	Kernel Procedures

L

LAN	Local Area Network
LCG	Local Communication Group
LEO	Low Earth Orbit
LL	Link Length
LMR	Lightweight Mobile Routing
LSR	Link State Routing
LSP	Link State Packets

M

MAC	Multiple Access Control
MIR	Minimum Impact Routing
MIR-VTP	MIR with Variable Transmit Power
MIR-VA	MIR with Variable-transmit-power plus Asymmetrical-routing
MIT	Massachusetts Institute of Technology
MF-TDMA	Medium Frequency - Time Division/Demand Multiple Access
MPR	Multipoint Relay
MPRS	Multipoint Relay Selector
MSSN	Multipoint Relay Selector Sequence Number

N

NDP	Neighbour Discovery Protocol
NFN	Next Forwarding Node
NNI	Network Node Interface / Network-to-Network Interface
NRN	Next Relaying Node
NTDR	Near-term Digital Radio
NTP	Network Termination Point

NWO	NetWork Operators
O	
OLSR	Optimised Link State Routing
OSI	Open System Interconnection
OSPF	Open Shortest Path First
OSS	Operational Support Systems
P	
PABX	Private Automatic Branch Exchange
PC	Power Control
PCF	Point Coordination Function
PHY	Physical Layer
POTS	Plain Old Telephone Service
PPTP	Point-to-Point Tunnelling Protocol
PR	Packet Relays
PRN	Packet Radio Network
PRD	Poison Random Distribution
PSTN	Public Switched Telephone Network
Q	
QoS	Quality of Service
R	
RA	Radio Communications Agency
RP	Reverse Path
RIP	Routing Information Protocol
RIPv2	Routing Information Protocol Version 2
RREQ	Route Request
RREP	Request Reply
RER	Route Error
RTS	Request To Send
RW	Route Weight

S

SC	System Capacity
SINR	Signal-to-Interference and Noise Ratio
SO	Satellite Operator
SOHO	Small Office Home Office
SP	Service Provider
SP	Shortest Path
SPR	Shortest Path Routing
SLA	Service Level Agreements
SME	Small to Medium Enterprises
SNR	Signal-to-Noise-Ratio
SNMP	Simple Network Management Protocol
SNWO	Satellite NetWork Operator
SR	Spatial Reuse
SRD	Spatial Reuse Distance
STD	State-Transition-Diagram
SURAN	Survivable Adaptive Radio Networks

T

TC	Topology Control
TCP/IP	Transmission Control Protocol over Internet Protocol
TDMA	Time Division Multiple Access
TDN	Total Disturbed Nodes
THC	Total Hops Counted
TINA	Telecom Information Networking Architecture
TORA	Temporally Ordered Routing Algorithm
TOM	Telecommunication Operation Map
TMF	Tele-Management Forum
TMN	Telecommunications Managed Network
TSIR	Time Sequenced Interference Region
TVRO	Television Receive-Only (Satellite Dish)

U

UDP/IP	User Datagram Protocol over Internet Protocol

UML	Unified Modelling Language

V

VTP	Variable Transmit Power
VPM	Virtual Process Manager
VSAT	Very Small Aperture Terminal
VSM	Virtual Service Manager
VPN	Virtual Private Network
VSAT	Very Small Aperture Terminal

W

WANET(s)	Wireless Ad hoc Network(s)
WN	Working Node (currently active and operating node)
WBAN	Wireless Body Area Networks
WPAN	Wireless Personal Area Networks
WLAN	Wireless Local Area Networks
WWAN	Wireless Wide Area Networks

X

XML	eXtensible Markup Language
x-DSL	A family of "Digital Subscriber Line" protocol

Z

ZRP	Zone Routing Protocol

Appendix III. References

1. Charles E. Perkins, *Ad Hoc Networking*. 2001: Addison-Wesley.
2. Ram. Ramanathan and Jason. Redi, *"A Brief Overview Of Ad Hoc Networks: Challenges And Directions"*, in *IEEE Communications Magazine*. May 2002. p. 20-22.
3. Li, J., et al. *Capacity of Ad Hoc Networks*. in *ACM SigMobile*. 2001.
4. B. M. Leiner, D. L. Nielson, and F. A. Tobagi, *"Issues in Packet Radio Network Design"*. Proceeding of the IEEE, Jan. 1987: p. 6~20.
5. NIST, *Mobile Ad Hoc Networks*,
   Tutorial: http://www.cs.tamu.edu/~vaidya/seminars.
6. NIST, J. Thompson, and Musenki, *"Mobile Ad Hoc Networks-Tutorial"*.
7. 11. Yu Ming Lu, *"Routing Protocols for Next Genereation Mobile Wireless Networks"*, in *Dept. of Electronics*. July 2003, Uni. of York: York.
8. Aleksi Penttinen, *"Research On Ad Hoc Networking: Current Activity And Future Directions"*.
9. E.M. Royer, S. Barbara, and C.K. Toh, *"A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Network"*. IEEE Personal Communication, April 1999: p. 46~55.
10. Martin Haenggi and Daniele Puccinelli, *"Routing in Ad Hoc Networks: A Case for Long Hops"*, in *IEEE Communication Magazine*. Oct 2005. p. 93-101.
11. M. Frodigh, P. Johansson, and P. Larsson, *"Wireless Ad Hoc Networking: The Art of Networking without a Network"*, in *No. 4*. 2000, Ericsson Review.
12. Edited By Mohammad. llyas, *The Handbook of Ad Hoc Wireless Networks*. 2001: CRC Press.
13. National Istitute of Standards and Technology (NIST), *"Wireless Ad Hoc Networks"*, Advanced Networking Technologies Division.
14. Madhavi. W. Subbarao, *"Ad hoc networking critical features and performance metrics"* Sep.1999, National Institutes of Standards and Technology (NIST).

15. Charles E. Perkin, *"Ad Hoc Networking"*. 2001: Addison-Wesley.

16. A. Agarwal and P.R.Kumar, *"Capacity bounds for ad hoc and hybrid wireless networks"*. ACM SIGCOMM Computer Communication Review, July 2004. **Volume** 34(Issue 3).

17. Yu Ming. Lu, David. Grace, and Paul. D. Mitchell. *"Capacity Evaluation of a Multi-hop Wireless Ad hoc Network Using Minimum Impact Routing"*. in *IEEE WiCOM 2006*. Sep. 2006. WuHan-China,.

18. Li. Jiandong, Haas. Z. J, and Sheng. M. *"Capacity Evaluation of Multi-Channel Multi-Hop Ad Hoc Networks"*. in *IEEE International Conference on Personal Wireless Communications (ICPWC)*. 2002.

19. Li. Jinyang, et al. *"Capacity of Ad Hoc Wireless Networks"*. in *ACM SigMobile*. 2001.

20. Gupta. P and Kumar.P. R, *"The Capacity of Wireless Networks"*. IEEE TRansactions on Information Theory, 2000. 46(2): p. 388-404.

21. Toumpis. S and Goldsmith. A, *"Capacity Regions for Wireless Ad Hoc Networks"*. IEEE Transaction on Wireless Communications, July 2003. No.4(2): p. 736-748.

22. A. S. Tanenbaum, *"Computer Networks"*. 3rd ed. 1996: Prentice Hall.

23. X.Y. Li, P.J. Wan, and O. Frieder. *"Coverage in Wireless Ad-hoc Sensor Networks"*. in *IEEE International Conference on Communications*. 2002.

24. Mike Youssefmir, et al., *"Criteria for Network Capacity"* May. 2003, IEEE 802 Executive Committee Study Group on Mobile Broadband Wireless Access.

25. T. Yoo, et al. *"Cross-layer Design for Video Streaming over Wireless Ad hoc Networks"*. in *International Workshop on Multimedia Signal Processing(MMSP-04)*. Sep. 2004. Siena, Italy.

26. National Science Foundation, *"Cross-Layer Design of Ad-hoc Wireless Networks for Real-Time Media"* June 2006.

27. F. Halsall, *"Data Communications, Computer Networks and Open Systems"*. 4th ed. 1996: Addison Wesley. 13~22.

28. D. Bertsekas and R. Gallager, *"Data Networks"*. 1987: Prentice Hall.
29. Mohammad. Ilyas, *"The Handbook of Ad Hoc Wireless Networks"*. 2001: CRC Press.
30. Yu Ming. Lu, *"HeliNet Task T4 internal Report on Investigation of IEEE 802.16"*. July 2002., Dept. of Electronics, Uni. of York: York..
31. Leonard. Kleinrock and John. Silvester. *"Optimum Transmission Radii for Packet Radio Networks or Why Six is a Magic Number"*. in *IEEE National Telecommunication Conference*. Dec. 1978.
32. Leonard. Kleinrock and Fouad A. Tobagi, *"Packet Switching in Radio Channels: Part I - Carrier Sense Multiple-Access Modes anad Their Thoughput-Delay Characteristics"*. IEEE Transactions on Communications, Dec. 1975 **com-23**(12): p. 1400-1416.
33. Yu Ming. Lu, et al. *"Performance Evaluation of Minimum Impact Routing for Multi-hop Wireless Ad Hoc Network"*. in *IEEE Wireless Personal Multimedia Communications (WPMC'04)*. Sep. 2004. Abano Terme, Padova, Italy.
34. G. Malkin, *"RIP Version 2---Carrying Additional Information"*, in *RFC 1388*. Jan. 1993, IETF.
35. Dong Kyun Kim, C. K. Toh, and Y. Choi. *"ROADMAP: A Robust ACK-driven Media Access Protocol for Mobile Ad Hoc Networks"*. in *IEEE ICCCN*. 2000. Las Vegas, USA.
36. Yu Ming. Lu and David.Grace, *"NEWCOM: First Report on Common Models Matching Project A Needs- Crosslayer Routing Design: Coping with Link Asymmetry and Minimum Impact Routing"*. Oct. 2005, EU Research Network - NEWCOM. p. 37-40.
37. Yu Ming. Lu, *"Routing Protocols for Next Genereation Mobile Wireless Networks"*, in *Dept. of Electronics*. July 2003, Uni. of York: York, UK.
38. X.Hong, K. Xu, and M. Gerla, *"Scalable routing protocols for mobile ad hoc networks"*. IEEE Network, Aug 2002. **16**(4).
39. A. Iwata, et al., *"Scalable Routing Strategies for Ad Hoc Wireless Networks"*. IEEE Journal on Selected Areas in Communications, 1999.

40. Mehmet.C.Vuran and Ian. F. Akyildiz, *"Spatial Correlation-Based Collaborative Medium Access Control in Wireless Sensor Networks".* IEEE/ACM Transactions on Networking, Apr. 2006. 14(2).

41. X. Guo, S. Roy, and W.S. Conner. *"Spatial reuse in wireless ad-hoc networks".* in *58th Vehicular Technology Conference-Fall (VTC)* 2003.

42. A. Fapojuwo, *"Wireless Communication System: Interferences and System Capacity in tutorial".* 2002, ENEL 519.29.

43. Theodorel. S. Rappaport, *"Wireless Communications: Principles and Practice".* 2nd ed. Dec. 2001.

44. IEEE Standard 802_11a Committee, *"Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, and IEEE Std 802.11a-1999 - High-speed Physical Layer in the 5 GHZ Band".* 1999.

45. Li, J.Y., et al. *Capacity of Ad Hoc Wireless Networks.* in *ACM/IEEE SIGMOBILE.* 2001.

46. H.Y. Hsieh and R.Sivakumar. *"IEEE 802.11 over multi-hop wireless networks: problems and new perspectives".* in *IEEE Vehicular Technology Conference (VTC).* 2002.

47. IEEE 802.3/4/5/6 standards committee, *IEEE standards for Local Area Network / Metropolitan Area Networks* http://www.ieee802.org/.

48. F.Ye, S.Yi, and B.Sikdar. *"Improving Spatial Reuse of IEEE 802.11 Based Ad Hoc Networks".* in *IEEE GLOBECOM.* 2003.

49. IEEE Standards Committee, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,* in *High-speed Physical Layer in the 5GHz Band.* 1999.

50. Jing Zhu, et al., *Adapting physical carrier sensing to maximize spatial reuse in 802.11 mesh networks.* Wireless Communications and Mobile Computing, Nov. 2004. 4(8): p. 933 - 946.

51. H.P. Fitzek, et al., *"Design and performance of an enhanced IEEE 802.11 MAC protocol for multihop coverage extension",* in *IEEE Wireless Communications.* 2003.

52. V. Kawadia and P.R.Kumar. *"Power control and clustering in ad hoc networks"*. in *INFOCOM*. 2003.

53. K. Tsudaka, et al., *"Power Control Routing for Multi Hop Wireless Ad Hoc Network"*. IEEE, May 2001.

54. X. Lin and I. Stojmenovic, *"Power-Aware Localized Routing in Wireless Networks"*. IEEE transaction on parallel and distributed systems, Oct. 2001. **12**(10).

55. S. Chakrabarti and A. Mishra, *"QoS Issues in Ad Hoc Wireless Networks"*, in *IEEE Communications Magazine*. Feb 2001. p. 142 ~148.

56. A.Iwata, et al., *Scalable Routing Strategies for Ad Hoc Wireless Networks* IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, 1999.

57. Lee, W.C.Y., *Spectrum Efficiency in Cellular.* IEEE Transactions on Vehicular Technology, 1989. **38**(2): p. 69-75.

58. Rajmohan Rajaraman, *"Topology control and routing in ad hoc networks: a survey"*. ACM SIGACT, June 2002. **33**: p. 60 - 73.

59. Paolo Santi, *"Topology control in wireless ad hoc and sensor networks "*. ACM Computing Surveys, June 2005. **37**: p. 164 - 194.

60. J. Wu, et al., *"On Calculating Power-Aware Connected Dominating Sets for Efficient Routing in Ad Hoc Wireless Networks"*. Journal of communications and networks, Korean Institute of Communication Sciences, March 2002. **4**(1).

61. IEEE 802. 3/4/5/6 standards committe, *"IEEE standards for Local Area Network / Metropolitan Area Networks"*.

62. G.Hiertz and J.Habetha, *"A new MAC protocol for a wireless multi-hop broadband system beyond IEEE 802.11"*. 2003: http://www.comnets.rwth-aachen.de/publications/Abstracts/HiHa_WWRF9-DRRP.html.

63. J.Li, et al., *"Performance evaluation of modified IEEE 802.11 MAC for multi-channel multi-hop ad hoc networks"*. Journal of Interconnection Networks, 2003. **4**: p. 345-359.

64. Shugong and T.Saadawi. *"Revealing TCP incompatibility problem in 802.11 based wireless muti-hop networks"*. in *IEEE Global Telecommunications Conference (GLOBECOM)*. 2001.

65. X.Shugong and T.Saadawi, *"Revealing the problems with 802.11 medium access control protocol in multi-hop wireless ad hoc networks"*. Computer Networks, 2002. 38(4): p. 531-548.

66. A. William and Shay, *"Understand Data Communications and Networks"*. 2004, Thomson Learning Asia Pte Ltd.

67. Ivan.Gojmerac, et al. *"Adaptive Multipath Routing for Dynamic Traffic Engineering"*. in *GLOBECOM*. 2003. San Francisco, USA.

68. C.E. Perkins and E. M. Royer. *"Ad-hoc On-demand Distance-Vector (AODV) Routing"*. in *Proceedings of the Second Annual IEEE Workshop on Mobile Computing Systems and Applications*. Feb. 1999.

69. Andrea J. Goldsmith and Stephen B. Wicker, *"Design Challenges for Energy-Constrained Ad Hoc Wireless Networks"*, in *IEEE Wireless Communications*. August 2002. p. 8-27.

70. Yu Ming. Lu. *"An Investigation of Routing Protocols for Broadband Wireless Communication Networks"*. in *IEEE PREP'02*. April 2002. Nottingham, UK.

71. J. Moy, *"OSPF Version 2"*, in *RFC 2178*. July 1997, IETF.

72. Li, X.-Y., P.-J. Wan, and O. Frieder. *Coverage in Wireless Ad-hoc Sensor Networks*. in *IEEE International Conference on Communications*. 2002.

73. E.M. Royer, P.M. Melliar.Smith, and L. E. Moser. *"An Analysis Of The Optimum Node Density for Ad hoc Mobile Networks"*. in *ICC 01*. 2001.

74. Robert. Sedgewick, *"Algorithms in C"*. 3rd ed. 2002: Addison Wesley. 265.

75. Thomas. H. Cormen, et al., *"Introduction to Algorithms"*. 2nd ed. 2001: MIT Press and McGraw-Hill.

76. E. W. Dijkstra, *"A Note on Two Problems in Connection with Graphs"*. Numerische Math, 1959: p. 269~271.

77. 1.Charles E. Perkin, *"Ad Hoc Networking"*. 2001: Addison-Wesley.

78. I. U. Javvin Technologies, *"Unicast and Unicast Routing"*.

79. Microsoft, *"Unicast Routing Overview."*

80. C.E. Perkins and P. Bhagwat, *Highly Dynamic Destination Sequence Distance Vector Routing (DSDV) for Mobile Computers.* ACM SIGCOMM'94, Computer Communications Review, , October 1994: p. 234~244.

81. P. Jacquet, P. Muhlethaler, and A. Laouiti, *"Optimised Link State Routing Protocol "*. July 2000, Internet-draft.

82. HIPERCOM, *"Multipoint Relays in OLSR."*

83. D. B. Johnson and D. A. Maltz, *"Dynamic Source Routing in Ad Hoc Wireless Networks"*. 1996: Kluwer Academic Publishers, Norwell, Mass.

84. C.E. Perkins and E. M. Royer. *Ad-hoc On-demand Distance-Vector (AODV) Routing.* in *Proceedings of the Second Annual IEEE Workshop on Mobile Computing Systems and Applications.* February 1999.

85. Vincent D. Park and M. Scott. Corson. *"A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks".* in *IEEE Conference on Computer Communications, INFOCOM'97.* April 1997. Kobe,Japan.

86. Chai.Keong.Toh, *"Associativity-Based Routing for Ad Hoc Mobile Networks".* Wireless Personal Communications: An International Journal archive, March 1997. 4: p. 103 - 139.

87. V. Park and M. S. Corson. *A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Network.* in *Proceedings of IEEE INFOCOM'97.* April 1997.

88. M. S. Corson and A. Ephremides, *"A Distributed Routing Algorithm for Mobile Wireless Network"* ACM/Baltzer Wireless Network Journal Feb. 1995. 1(1): p. 61-82.

89. E. Gafni and D. Bertsekas, *"Distributed Algorithms for Generating Loop-free Routes in Networks with Frequently Changing Topology"* IEEE Transactions on Communications, January 1981. 29(1): p. 11~15.

90. A. E. M. Weber (Bates). *"A Simulated Performance Study of Some Distributed Routing Algorithms for Mobile Radio Networks"* in *Conference on Information Sciences and Systems (CISS'89).* March 1983.

91. C. C. Chiang. *"Routing in Clustered Multihop, Mobile Wireless Networks with Fading Channel"* in *IEEE SICON'97*. April 1997.

92. Z. J. Haas and M. R. Pearlman, *"Internet draft:The Zone Routing Protocol for Ad hoc Networks"*. Nov 1997, IETF,Mobile Ad hoc Networking group.

93. A. Ephremides, J. E. Wieselthier, and D. J. Baker, *"A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signalling"*. Proceeding of IEEE, January 1987. 75(1): p. 56-73.

94. D. J. Baker and A. Ephremides, *The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm.* IEEE Transactions on Communications COM-29 (11), November 1981: p. 1694-1701.

95. M. Gerla and J. T. C. Tsai, *Multicluster, Mobile, Multimedia Radio Network.* Wireless Networks 1(3), October 1995: p. 255~265.

96. Deepak Ganesany, et al., *"Complex Behavior at Scale: An Experimental Study of Low-Power Wireless Sensor Networks"*, T.R.U.C.-T. UCLA Computer Science, Editor. 2003.

97. Dong Kyun Kim, C.K.Toh, and Yanghee Choi. *"On Supporting Link Asymmetry in Mobile Ad Hoc Networks"*. in *IEEE Symposium on Ad Hoc Mobile Wireless Networks in conjunction with IEEE GLOBECOM 2001*. Nov. 2001.

98. DongKyun Kim, C. K. Toh, and Y. Choi. *ROADMAP: A Robust ACK-driven Media Access Protocol for Mobile Ad Hoc Networks.* in *IEEE ICCCN, 2000.* 2000. Las Vegas, USA.

99. Jong-Hoon Lee, et al. *"An Asymmetric Routing Protocol Using Detour in Mobile Ad Hoc Networks"*. in *ICOIN' 02*. 2002. Berlin Heidelberg, Germany.

100. Thomas T. Fuhrmann, *"Protocol Independent Multicast and Asymmetric Routing"*. Jan 2000, Praktische Informatik IV, Technical Report.

101. Leonard. Kleinrock, *"Queueing Systems - Volume I: Theory & Volume II: Computer Applications"*. 1975: John Wiley & Sons.

102. OPNET, *"OPNET Modeler - Modeling Concept - Chapter 1 Modeling Overview"*, in *OPNET On-line documents.*

103. Kostas Pentikousis, et al., *"Active Goodput Measurements from a Public 3G/UMTS Network"*. Sep. 2005. **IEEE COMMUNICATIONS LETTERS VOL 9, NO. 9.**

104. ITU-D Study Group 2, *"Teletraffic Engineering"*. Jan. 2005, Geneva.

105. André B. Bondi. *"Characteristics of scalability and their impact on performance"*. in *2nd International Workshop on Software and Performance.* 2000. Ottawa, Ontario, Canada,.

106. Fouad A. Tobagi and Leonard. Kleinrock, *"Packet Switching in Radio Channels: Part II - Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution"*. IEEE Transactions on Communications, Dec. 1975 **com-23**(12): p. 1417-1433.

107. Yu Ming. Lu, *"Resource Allocation Methods and Network Protocols for HeliNet "*. July 2002, EU Research Project-Framework 6-HeliNet-Task 4 internal Report HE-064. p. 117~120.

108. Hoi.Sheung. Wilson. So, Giang. Nguyen, and Jean. Walrand. *Practical synchronization techniques for multi-channel MAC* in *12th annual international conference on Mobile computing and networking, Los Angeles, CA, USA* 2006