The Design of Ad Hoc Networks
with Minimum Power
and Maximum Battery Life

Danish Khan

PhD 2009

Abstract

Multi-hop wireless ad hoc networks consist of terminals that can communicate without the support of fixed infrastructure. Nodes may communicate directly from source to destination or can use other nodes in the network as relays to facilitate a path from source to destination. These networks can be rapidly deployed and are therefore well suited to emergency service applications where fixed infrastructure has become unavailable or in situations where a temporary network is required.

The essence of this type of network is the agreed co-operation between users and the underlying principle that each user is willing to make itself available as a relay for the overall benefit of the network group. In most cases the terminals in these networks are battery powered and so, to maximise the lifetime of the network, the power consumption of each node needs to be managed.

This thesis studies the optimum design of a multi-hop ad hoc network. Each layer is analysed and cross layering is considered where it is able to improve the performance. Four routing strategies for managing node usage are investigated; a minimum power routing scheme, a minimum power routing with a battery charge threshold scheme, a residual battery charge scheme and a proposed minimum power routing/maximum battery lifetime scheme. A network model has been developed to evaluate these schemes and the results show that a network lifetime (defined as the time until the first node reaches zero battery charge) of 21 hours can be obtained using the proposed routing scheme which represents an improvement of 5% over the power aware routing scheme and the residual battery charge scheme, 31% over the minimum power routing with a battery charge threshold scheme, and 133% over the minimum power routing scheme.

Space, frequency and time division multiple access schemes are analysed for supporting multiple simultaneous transmissions in the network. Space division multiplexing allows multiple access without affecting the bandwidth or data rate, but five to nine simultaneous routes can be supported. A time division scheme is considered the best solution when guaranteed access is required by all nodes, but this reduces the maximum bit rate per user.

The throughput per unit time in a multi-hop route using a single frequency channel varies inversely with the number of hops. A novel cross layer scheme is proposed that selects the modulation order to match the number of hops in a route to maximise the throughput per unit time. Simulation results for this scheme show that this can improve the throughput on 52% of the routes using a proposed routing scheme, but the extra power required for the higher order modulation reduces the network lifetime by 14%.

A total network design solution is presented, including both the transmission and signalling subsystems that shows how the novel routing and cross layer features proposed in the thesis can be implemented.

Acknowledgements

I would like to thank my supervisor Dr Peter Ball for his continuous support throughout my research. The door to Dr Ball's office was always open whenever I had a question about my research or writing. He is an outstanding researcher who can constructively criticise research and an excellent supervisor who can bring the best out from his students. He consistently allowed this research to be my own work, but steered and guided me in the right direction. I am thankful to my second supervisor Dr Geoff Childs for his feedback and comments at various stages which have been significantly useful in shaping the thesis up to the completion. The good advice and support of my Postgraduate Tutor, Dr Khaled Hayatleh, has been invaluable on both an academic and a personal level, for which I am extremely grateful.

Foremost, I am thankful to my parents who covered financial matters of my research; my brothers Kamran & Irfan, and sister-in-law Hina, who provided emotional support, motivation and took great care of me in many aspects. I am also thankful to my uncle and aunt Usman & Mehmooda for their help during my research work.

There are number of people in my every day life who have enriched my professional life in various ways. I am thankful to my colleagues Hui, Mei, and Hall Staff Members especially Robin and Lamin for their friendship and moral support.

I would like to thank my fiancée Nilay Dönmez for her patience, sacrifice and continuous encouragement to support me mentally from abroad, and especially the time she spared for me from her own professional commitments which was so precious. I am thankful to my future father & mother-in-law and sister-in-law Ayşegül, for understanding my long-term research commitment.

Finally, and most importantly, I would like to thank the Almighty God, for it is under His grace that we live, learn and flourish.

Table of Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

2G	Second Generation
3G	Third Generation
A	Ampere
ABR	Associativity Based Routing
ACK	Acknowledgement
AODV	Ad hoc On-demand Distance Vector
AP	Access Point
ARPANet	Advanced Research Projects Agency Network
BER	Bit Error Rate
BFA	Bellman-Ford Algorithm
BPSK	Binary Phase Shift Keying
BQ	Broadcast Query
C/I	Carrier to Interference
CDMA	Code Division Multiple Access
CGSR	Clusterhead Gateway Switch Routing
CMMBCR	Conditional Maximum Minimum Battery Capacity Routing
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
CTS	Clear to Send
DAA	Dynamic Adaptive Acknowledgement
DARPA	Defense Advanced Research Projects Agency
dB	Decibel
DRP	Dynamic Routing Protocol
DSDV	Destination Sequence Distance Vector
DSR	Destination Source Routing
DSSS	Direct Sequence Spread Spectrum
E_b/N_o	Signal energy per bit to noise power spectral density ratio
ECN	Explicit Congestion Notification
EDGE	Enhanced Data rates for GSM Evolution
EGPRS	Enhanced General Packet Radio Service
ELN	Explicit Loss Notification
FDMA	Frequency Division Multiple Access

FSR	Fisheye State Routing
FTP	File Transfer Protocol
GRM	Ground Reflection Model
GSM	Groupe Spécial Mobile
HTTP	Hypertext Transfer Protocol
Hz	Hertz
IARP	Intrazone Routing Protocol
ICMP	Internet Control Messaging Protocol
IEEE	Institute of Electrical and Electronics Engineers
IERP	Interzone Routing Protocol
IETF	Internet Engineering Task Force
IN	Intermediate Node
IP	Internet Protocol
IPv4	Internet Protocol version 4.0
IPv6	Internet Protocol version 6.0
ISM	Industrial, Scientific and Medical
LAN	Local Area Network
LAN-MAR	Landmark Routing
LOS	Line of Sight
MAC	Medium Access Control
mA-hrs	milli-Ampere Hours
MANET	Mobile Ad Hoc Network
MB	Megabyte
Mbps	Megabits per second
MPR	Minimum Power Routing
MPR/MBL	Minimum Power Routing/Maximum Battery Lifetime
OLSR	Optimized Link State Routing
PHY	Physical Layer
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RERR	Route Error
RFC	Request for Comments

RREP	Route Reply
RREQ	Route Request
RSS	Received Signal Strength
RTS	Request to Send
Rx	Receiver
SDMA	Spatial Division Multiple Access
SINR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
SRD	Source, Relay, Destination
SRP	Static Routing Protocol
SSR	Signal Stability Routing
SST	Signal Stability Table
TDMA	Time Division Multiple Access
TETRA	Terrestrial Trunked Radio
TTL	Time To Live
Tx	Transmitter
UDP	User Datagram Protocol
U-NII	Unlicensed National Information Infrastructure
W	Watt
WLAN	Wireless Local Area Network
WRP	Wireless Routing Protocol
ZRP	Zone Routing Protocol

1. Introduction

1.1 Mobile ad hoc networks - history

1.1.1 Wireless packet networks

Packet switched technology, first demonstrated by Defense Advanced Research Projects Agency (DARPA) called Advanced Research Projects Agency Network (ARPANet) in the 1960s, provided great promise for bandwidth sharing among multiple users and provides adaptive routing in response to changing network conditions and user demands. Inspired by the advantages of packet switching, in 1972 DARPA initiated a research project to develop and demonstrate a Packet Radio Network (PRNet) [1]. The main intention of the PRNet was to provide an efficient means of sharing a broadcast radio channel with minimal central control, coping with changing and incomplete connectivity.

A modified version of the ALOHA protocol was used as a medium access in PRNet. ALOHA was originally developed in 1970 at the University of Hawaii, in order to link remote campuses of the University over Hawaiian Islands [2]. In addition, PRNet used a multi hop store-and-forward routing technique to provide an end to end complete radio path.

The core ideas of the PRNet were applied in a number of environments such as terrestrial and airborne, narrowband and wideband, amateur radio and satellite. All of these networks shared a number of characteristics of a mobile ad hoc network. In parallel with PRNet, another DARPA project, packet switching over satellites (SatNet) [3] was created, and from that point, there was a need to merge ARPANet, PRNet and SatNet together. This was the origin of the Internet program launched in 1974 [4].

Survivable Radio Networks (SURANs) were developed by DARPA in 1983 to address open issues in PRNet such as network scalability, security, processing capability and energy management. The aim of SURANs was to develop network algorithms that scale to tens of thousands of nodes to provide robust security to protect packet radio network from security attacks, and to use a low-cost, low-power radio that would support sophisticated packet radio protocols [5]. The effort resulted in the design of Low-cost

Packet Radio (LPR technology in 1987 [6]), a digitally controlled Direct Sequence Spread Spectrum (DSSS) radio with an integrated microprocessor-based packet switch.

As interest in PRNet grew, particularly in the United States Army for spread-spectrum radios, DARPA and the Army worked together in a series of testbeds in the 1980s, to demonstrate how ground command and control can be improved through the use of PRNet and Internet technologies [7].

In the late 1980s and early 1990s, mobile computing become more affordable in the form of laptops, notebooks and personal digital assistants, and at the same time, hardware and software, especially open source software, became widely available. In addition, the widespread implementation of Internet and Web technology provided an incentive to the commercial and defense sectors to extend the global information infrastructure into the mobile wireless environment. The Institute of Electrical and Electronics Engineers (IEEE) 802 project which relates to computer communication networks, established the 802.11 working group to standardise techniques and technologies to be used for wireless local area networks (WLANs). The first IEEE Workshop on Wireless Local Area Networks (WLANs) began in 1991. The experts were working in parallel to address both infrastructure and infrastructureless needs.

In 1994, DARPA initiated the Global Mobile (GloMo) Information System [8] which aimed to provide multimedia connectivity between wireless handheld devices.

Tactical Internet (TI), implemented by the U.S. Army in 1997, is by far the largest scale implementation of a mobile wireless multi hop packet radio network [5]. Nodes consist of both vehicular and man-packed radios using position location reporting systems and ground to airborne radio systems, running modified Internet protocols. It uses a direct-sequence spread spectrum time division multiple access radio capable of transmitting data at tens of kilobits per second.

In the same year, the 802.11 working group (WG) released the first WLAN standard which is IEEE 802.11 for both infrastructure and infrastructureless networks.

In 1999, the Extending the Littoral Battlespace Advanced Concept Technology (ELB ACTD) demonstration demonstrated the feasibility of communications from ships at sea to marines on land via an aerial relay [9]. A commercial WLAN product from Lucent's WaveLAN was used with VRC-99A, a direct sequence spread spectrum radio from the DARPA. WaveLAN was used to connect to an access point on a terrestrial or airbone relay, and VRC-99As were used as the mobile backbone to connect the terrestrial or airborne routers. The number of nodes was 20 and ELB ACTD successfully demonstrated the use of aerial relays for connecting users beyond line of sight (LOS).

1.1.2 Ad hoc networks

Internet Engineering Task Force (IETF)

When ARPANet was set up, Department of Defense (DoD) created an informal committee to oversee it. In 1983, the committee was renamed as IAB (Internet Activities Board), and was later changed to Internet Architecture Board. In 1989, the IAB was reorganized and researchers were moved to the Internet Research Task Force IRTF and the Internet Engineering Task Force (IETF). The reason for the split was to have IRTF concentrate on long-term research and IETF to deal with short-term issues. The IETF was divided into groups, each with a specific problem to solve [10].

One of the IETF working groups is Mobile Ad hoc Network (MANET). The purpose of this group is to standardize Internet Protocol (IP) routing protocol functionality suitable for static and dynamic ad hoc networks. This group has produced several drafts for routing protocols [11].

802.11 Wireless LAN standard

After the release of the first WLAN standard from IEEE 802.11 WG in 1997, several task groups such as 802.11a, 802.11b, 802.11g, 802.11e, etc., were formed. Each of these task groups looked at a different aspect of WLANs such as the physical layer, medium access, security, quality of service, handover, and interoperability between WLAN technologies. Amendments of the 802.11 standard from each task group are available separately, however in 2007, the 802.11 standard was revised to incorporate all previously approved amendments from the separate task groups into a single IEEE 802.11 standard for WLAN [12].

The original 802.11 standard defines an ad hoc network as an infrastructureless network using point to point links. Multi hop ad hoc network was not included originally in the 802.11 standard. The growing interest of wireless LAN applications has encouraged the commercial development of wireless mesh networks. As a result, several companies and manufacturers started selling proprietary solutions for wireless mesh networks for both indoor and outdoor applications, using the 802.11 standard [13]. However, these solutions were using proprietary routing and Medium Access Control (MAC) protocols and software which make prevented interworking between different systems. There was therefore a need for an open standard to bring down the cost of equipment and ensure interoperability. For this reason, in 2004, the 802.11 Study Group for Mesh Networking was created. In the same year, the 802.11s Task Group was formed [14]. The aim of the task group is to extend the IEEE 802.11 architecture and protocol to support mesh networking and to enable automatic topology learning and path configuration to create an ad hoc network.

1.2 Design considerations for ad hoc networks

Wireless ad hoc networks consist of wireless nodes which can dynamically self organize into arbitrary and temporary ad hoc network topologies. They have the feature that they can be rapidly deployed allowing nodes to connect with no fixed communication infrastructure. Ad hoc networks can be formed in different ways depending on the application. For example, nodes can communicate with each other point to point, or nodes can form a link to the destination node with the help of relays. The latter is called a multi hop ad hoc network. Relays in the ad hoc network can be used to extend the range over which nodes can communicate or they can be exploited to circumnavigate obstacles that would form a barrier for a single hop link. These features make ad hoc networks attractive for applications such as wireless personal area networks, law enforcement operations, search and rescue operations and wireless sensor networks. However, these benefits introduce some design constraints as outlined below.

1.2.1 Medium access

The IEEE 802.11 MAC protocol is the most commonly used MAC protocol in ad hoc networks. This protocol generally uses carrier sense multiple access with collision avoidance (CSMA/CA) with extensions to allow for the exchange of Request to

Send/Clear to Send (RTS/CTS) control packets [12] between the source and the destination node. These control packets are transmitted before the data transmission to prevent other potentially interfering nodes from transmitting. However this process prevents concurrent transmissions over the footprint of the control signals.

1.2.2 Routing

Routing protocols for wired networks have not been designed to support self-forming, and self-configuring networks. The topology of an ad hoc network may be static or dynamic depending on the scenario. Wired routing protocols do not adapt according to the network condition. For example, if links in the network are static, then sending routing updates when there is a change in the link is effective in maintaining the best route. However, if the nodes in the network are moving, then routing tables must be updated frequently to track the change in topology and this required additional signalling.

1.2.3 Energy consumption

Nodes in an ad hoc network may be directly connected or a multi hop route may be formed with the support of relay nodes if the destination is not in range or blocked due to buildings. In a mobile ad hoc network, the nodes are often battery powered, and if a node is continuously selected as a relay, then the node's battery may be depleted rapidly and then nodes will not then be able to participate in the network [15]. The selection criteria for relay nodes should therefore taken into account not only minimising power consumption but also maximises the time over which all nodes can communicate.

1.3 Reference model for ad hoc networks

Most data networks are based on the Transmission Control Protocol/Internet Protocol (TCP/IP) model, which adopts a layered approach to support the functions required for data transmission [16]. A strictly layered approach to data transmission has the advantage that each layer can be designed independently of every other layer which facilitates interfacing between the equipment from different manufacturers. This approach has been very successfully in fixed line applications including LANs and the Internet.

Over the last 15 years, there has been a tremendous growth in wireless networks. These differ from fixed line networks in a number of ways:

- Mobility: Users can move freely whilst using network services. As the user moves, the transmission path varies, which can lead to a variation in the signal strength.
- Variable propagation: Attenuation can vary due to weather conditions and due to variable obstructions such as trees and vehicles.
- Battery powered handsets: Battery operated wireless devices have a finite energy source and a recharging facility is not always available.
- Security: The air medium for communication is vulnerable for eavesdropping which creates potential security hazards.

The layered TCP/IP approach is not well suited to wireless networks. For example, in a wired network, packet loss is generally due to congestion and the response is to re-transmit the lost packets. However in a wireless network, packet loss may not only be due to congestion but it could also be due to fading or attenuation of the signal over the wireless medium. If the cause is fading or attenuation, retransmissions is not an appropriate solution as the packets will also be lost on the retransmissions. For the wireless system to take the appropriate action, information from the physical layer about the received signal strength needs to be passed to the transport layer so that the transport layer can distinguish between packet loss due to congestion and signal variation. This form of communication between the layers is known as cross layer design [17-19].

1.4 Research motivation

Many of the existing wireless networks such as cellular Second Generation (2G), Third Generation (3G) and beyond, satellite networks and dedicated public safety networks such as Terrestrial Trunked Radio (TETRA) use a fixed communication infrastructure (tower or masts). However, if the infrastructure is damaged or out of service for any reason, perhaps through a natural disaster such as an earthquake or a flood, then the communication network will fail.

In such circumstances, wireless ad hoc networks can be used to provide temporary communications to support the rescue services, as they can be rapidly deployed and do not rely on any infrastructure [20]. Moreover the ability to use multi-hop routes increases the likelihood of communication between nodes and can extend the overall communication range. Key factors in the design of this type of network, therefore, are maximising the lifetime and maintaining the required acceptable quality of service so that communication between all nodes can be sustained for as long as possible. It is the aim of this research to explore ways in which the lifetime and performance of ad hoc networks can be improved, such that in mission critical applications it may ultimately lead to lives being saved.

1.5 Research objectives

The overall aim of this research is to study the design of multi-hop wireless ad hoc networks and to investigate techniques for improving their operation and performance. More specifically the work will address the following:

i. Investigate techniques for improving the lifetime of ad hoc networks with battery powered terminals.
ii. Study methods for maximising the throughput in multi-hop ad hoc networks and making the throughput independent of the route selection.
iii. Consider the most appropriate architecture for an ad hoc network, considering both layered and cross layered approaches.
iv. Identify a network solution that can implement the schemes proposed in this research.
v. Develop a network model to characterise and evaluate the performance of the proposed scheme.

1.6 Thesis structure

Chapter 2 is an introduction to ad hoc networks, and includes a discussion about the medium access, network and transport layer design issues.

Chapter 3 discusses the limitations of using a layered protocol architecture in wireless networks, and investigates whether the use of cross layer interactions can improve the system network performance.

Chapter 4 addresses the issue of power consumption in an ad hoc network. Minimum power routing schemes are reviewed and, more specifically, schemes for maximising the network lifetime are considered. A model is developed to predict the evolution of the battery charge as a function of time for each node in a network. This is used to investigate the use of battery thresholds to balance the use of all the nodes in the network and maximise the time when all nodes are available for communication. A routing scheme based on residual battery charge is analyzed. A routing scheme is proposed in which minimum power and residual battery charge schemes are combined in order to optimise the balance between the minimum power routes and the maximum network lifetime.

Chapter 5 presents a total system solution for an ad hoc network that is able to support the schemes, described in chapter 4, for extending the network lifetime. A solution is described for both the transmission path and the associated signalling needed to distribute key parameters between the network nodes. The propagation model is explained and the radio frequency parameters are defined together with the modulation schemes and the data transmission rate. The design uses parameters for commercially available components.

Chapter 6 describes a model that has been developed to analyse the performance of the ad hoc network with the features described in chapters 4 and 5. Simulation experiments carried out with the model to evaluate the network performance. The results are presented and analysed and compared to previously reported results.

Chapter 7 discusses multiple access approaches and spatial re-use is analysed with the help of network model. This chapter also consider schemes for managing throughput and it includes a mechanism for implementing the system using cross layer design.

In chapter 8, simulation experiments carried out with the model to evaluate the network performance when throughput is maintained.

Chapter 9 presents the conclusions of the work in this thesis. Original contributions are highlighted and areas for future work are discussed.

2. Mobile Ad Hoc Networks

2.1 Introduction

Wireless networks can be divided into two broad categories: infrastructure-based networks and infrastructure-less networks. Infrastructure-based wireless networks are based on a fixed network with wireless access points to communicate with mobile terminals. This type of network is used in cellular radio systems and wireless LANs. In infrastructureless wireless networks, each mobile wireless terminal can establish a wireless connection to another wireless terminal either directly or by using other nodes as relay nodes. These networks are referred to as mobile ad hoc networks.

Multi-hop ad hoc wireless networks have several advantages. Firstly, because nodes in an ad hoc network can configure themselves and form a network without the aid of any fixed infrastructure, they can be deployed where quick formation of a network is required, for example, in a battle field, in rescue operations or in conferences. Secondly these networks are very flexible as they have the ability to be able to form a mesh network in which there can be several different routes from a source node to a destination node. This gives resilience against the failure of a node or link obstruction. Multi-hop networks also have the advantage that they can enable nodes to communicate over a larger area than would be possible with single hop networks. Furthermore, where there is a high attenuation between the source and destination, a multi-hop route may be able to reduce the end to end power consumption by finding a lower loss route.

The challenges that face multi-hop ad hoc networks include the need for signal processing at each node to set up and maintain paths across the network which can result in an increase in node complexity. The following sections outline the key design issues for mobile ad hoc networks.

2.2 Design issues for multi hop ad hoc networks

2.2.1 Medium access control in mobile ad hoc networks

Radio is inherently a broadcast medium where each transmission is heard by several receivers. Nodes in an ad hoc network use radio waves which propagate in all directions when an omni directional antenna is used. Other nodes in the network can sense such transmission, as shown in figure 2.1. In order to access the medium, joint coordination is required where one node is allowed to transmit and other node waits for the medium to access.

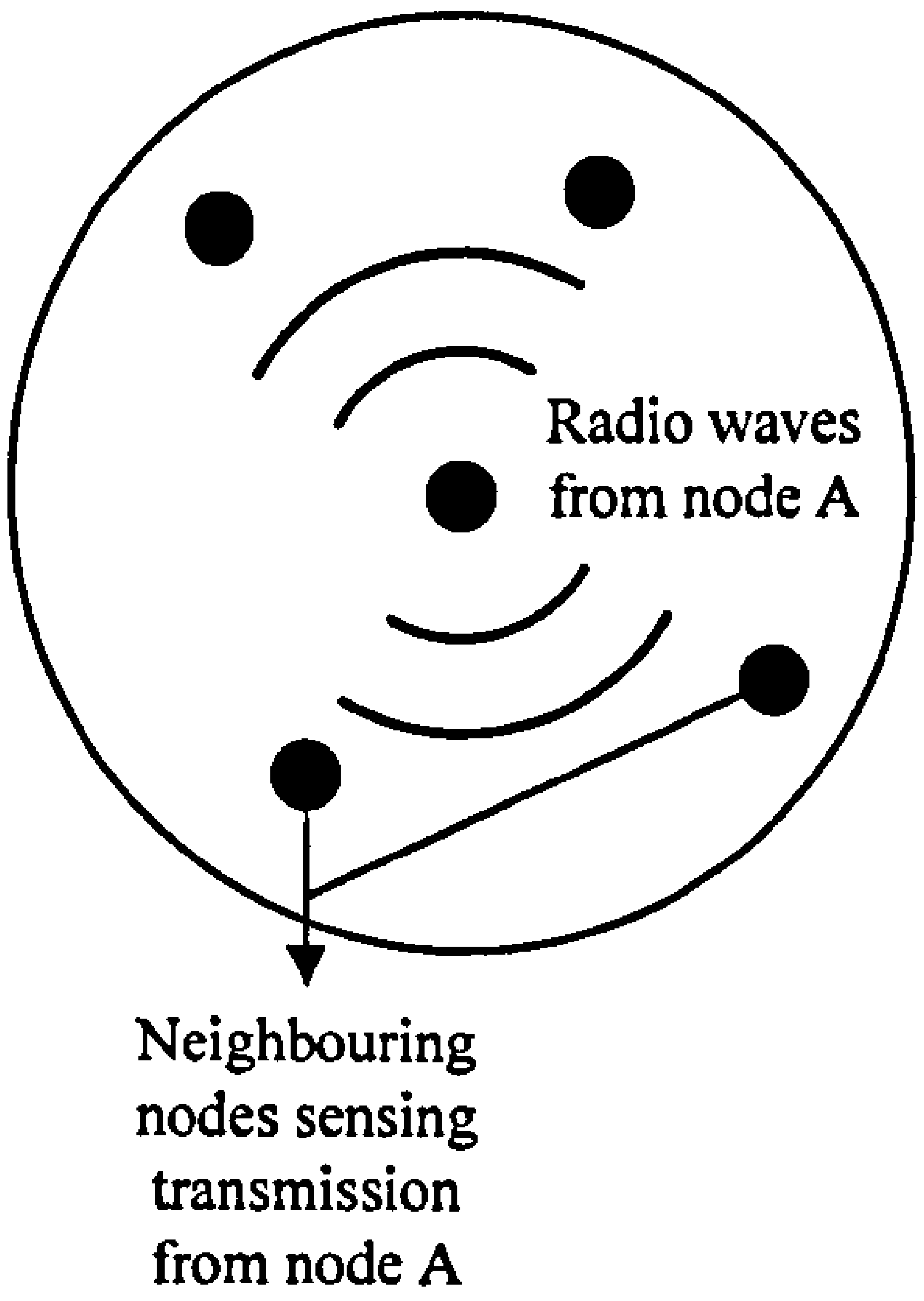


Figure 2.1 Broadcast wireless medium

2.2.1.1 Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA)

In order to avoid interference, medium access needs to be managed. One scheme for multiple access commonly used in wireless LANs is CSMA/CA [16]. In this scheme any node that wants to transmit, first senses the medium. If the medium is busy (for example, another node is transmitting), then the node will defer its transmission until a later time. If the medium is free, then the node is allowed to transmit. Unfortunately this scheme does not work well in a mobile ad hoc network due to the problem of hidden and exposed nodes [16].

2.2.1.2 CSMA/CA with RTS/CTS

Various MAC schemes have been proposed for ad hoc networks [21-26] to resolve the hidden terminal problems and improve channel utilization. These MAC protocols involve the use of a RTS and CTS frames. Such an approach has been accepted by the IEEE 802.11 Wireless LAN standard [12]. Data and control frames are transmitted at a fixed maximum power in order to notify other nodes about an on-going communication.

2.2.1.3 Transmission power consideration

Transmitting at the maximum power maximises the range of communication, but it also maximises the area over which other nodes are prevented from transmitting. For example, in figure 2.2, node C can sense data transmission from node A if it transmits at the maximum power, but if nodes A and B use the minimum power to communicate, then nodes C and D can also initiate concurrent transmission. Secondly when a source node transmits at the maximum power, the power may be larger than the minimum signal needed to achieve the required signal to interference ratio (SIR), which wastes the node's battery and reduces lifetime. Fixed power data and control packet transmission can reduce the probability of being able to support simultaneous transmissions in the network (decrease network throughput) and use battery inefficiently.

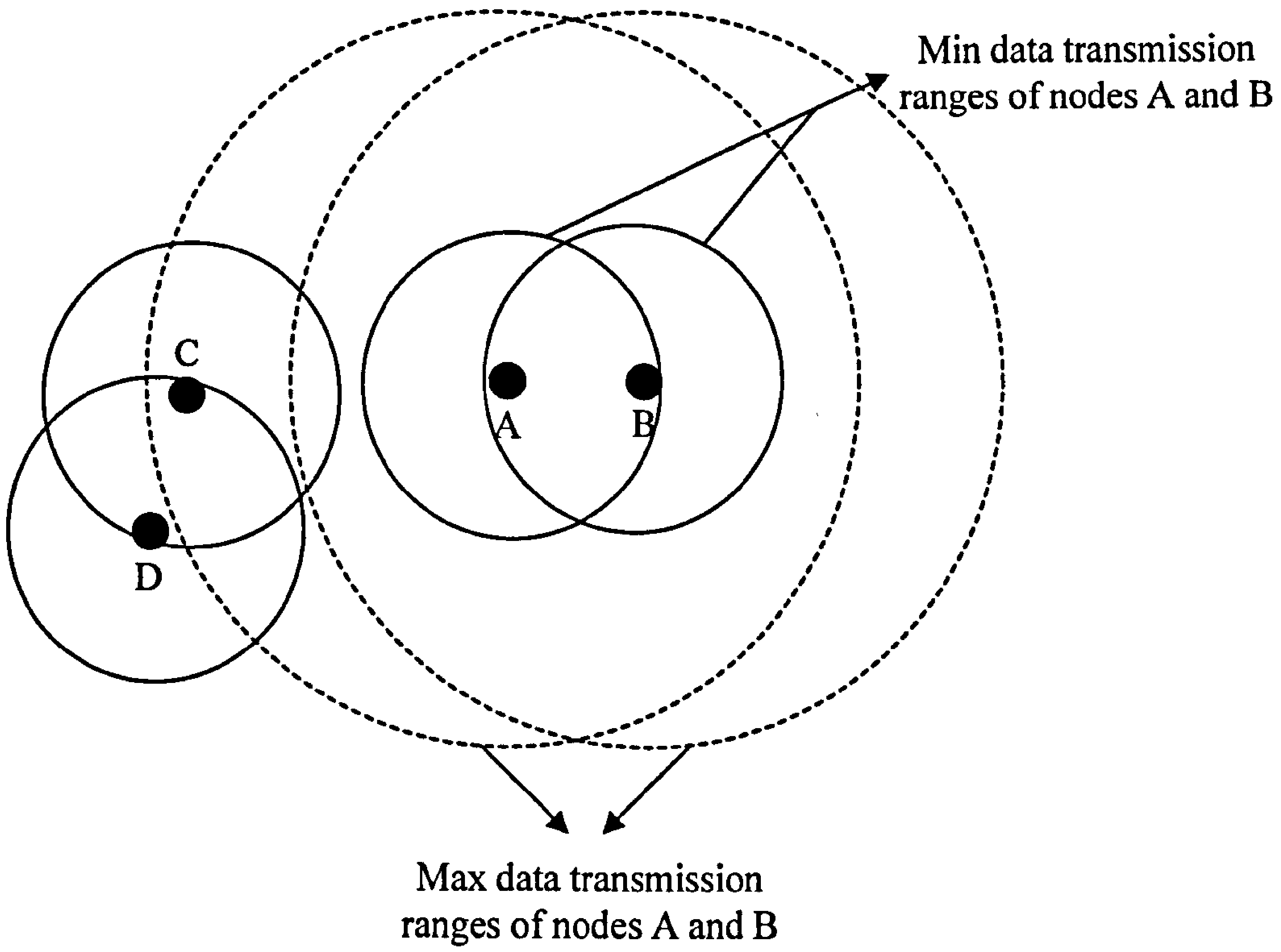


Figure 2.2 Simultaneous transmissions using power control approach

Another approach is to use a variable transmit power. In the literature, various approaches are proposed (as a modification to the IEEE 802.11) to control power levels of the data and control signals in order to improve the channel utilization and minimise the battery consumption. Transmit power can be saved by transmitting the minimum power required to reach the receiver. However, this scheme can create a collision with an on-going transmission. For example, in figure 2.3, node C is receiving data from node D at the minimum power. If node A wants to communicate with node B, node A does not sense any on-going communication and sends a RTS packet to node B. Node B also senses the free medium, and replies with CTS. Because nodes C and D are using minimum power to communicate, neighbouring nodes A and B are not aware of any data transmission in the neighbourhood and hence node A transmission causes interference at node C. Furthermore, if nodes A and B started communication earlier than nodes C and D, then nodes C and D will not able to start a simultaneous communication.

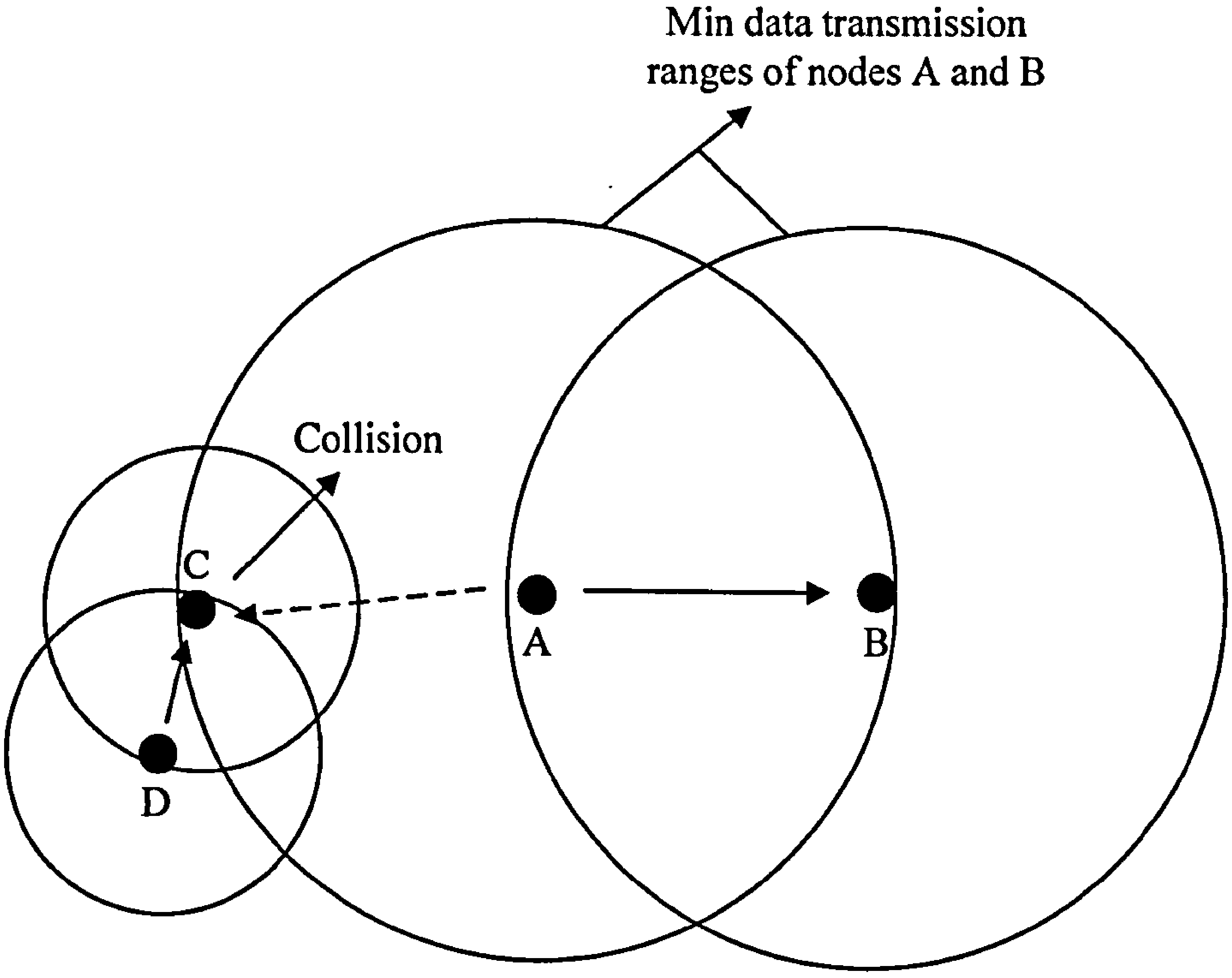


Figure 2.3 Hidden node problem in the minimum transmit power approach

One solution to this problem is to transmit control signals (RTS/CTS) at the maximum power, and DATA/Acknowledgement (ACK) at the minimum power necessary for the nodes to communicate [26-28]. When the control packets are transmitted at the maximum power, distant nodes can hear this packet and are prevented from transmitting. For example, in figure 2.4, the RTS/CTS transmissions between node C and node D are transmitted at the maximum power and data is transmitted at the minimum transmission power. The neighbouring nodes such as node A will hear control signals and will not initiate a simultaneous communication with node D. Hence collision is prevented at node C as illustrated in figure 2.3.

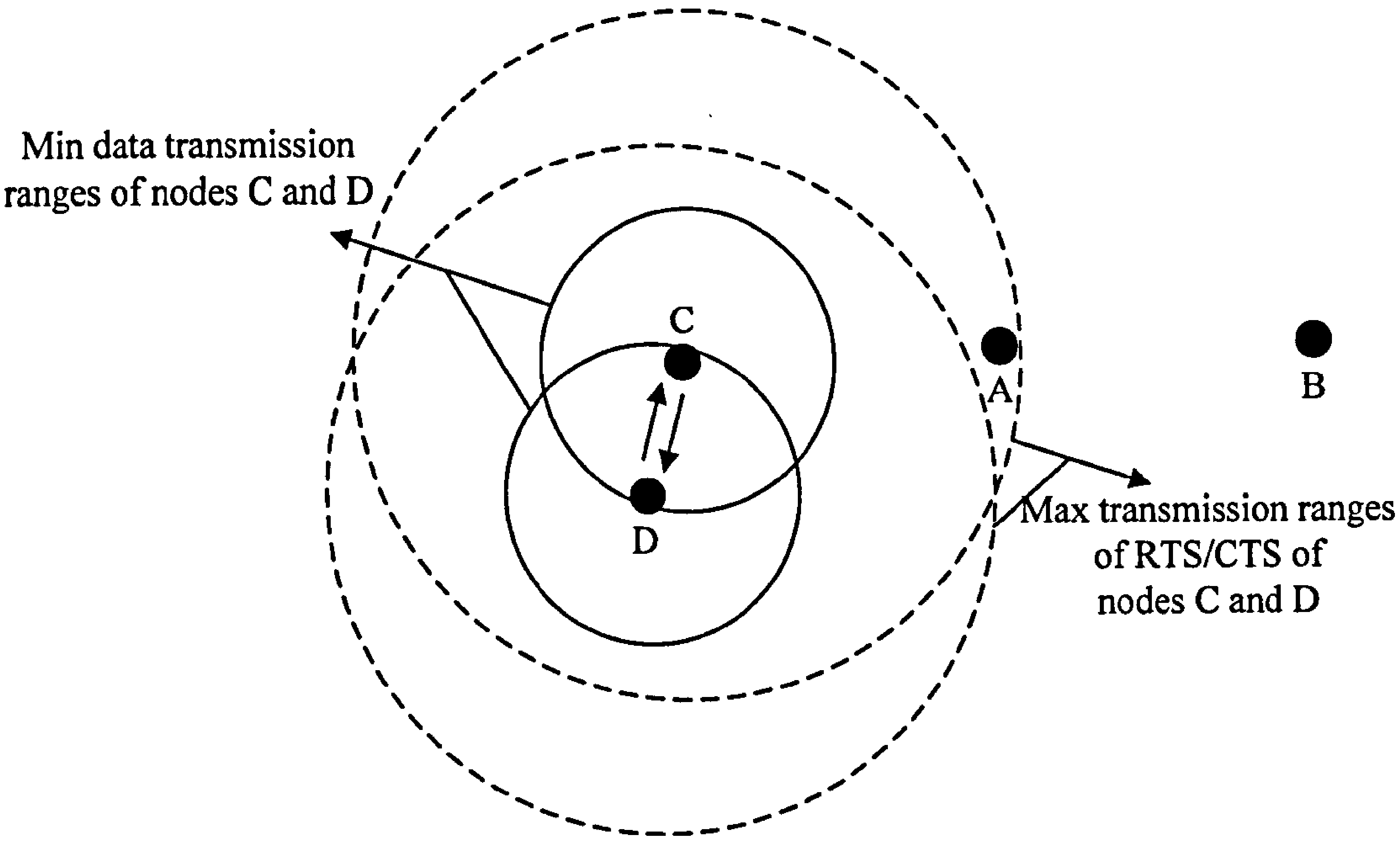


Figure 2.4 Minimum data transmission and maximum RTS/CTS ranges

However transmitting control signals at the maximum power and DATA/ACK at the minimum power prevents neighbouring nodes to initiate simultaneous communication. For example, in figure 2.5, node A and node B data transmission will not interfere with on-going data transmission between node C and node D. However due to the maximum power transmission of control signals between nodes C and D, node A will hear these control signals which prevents simultaneous communication with node B.

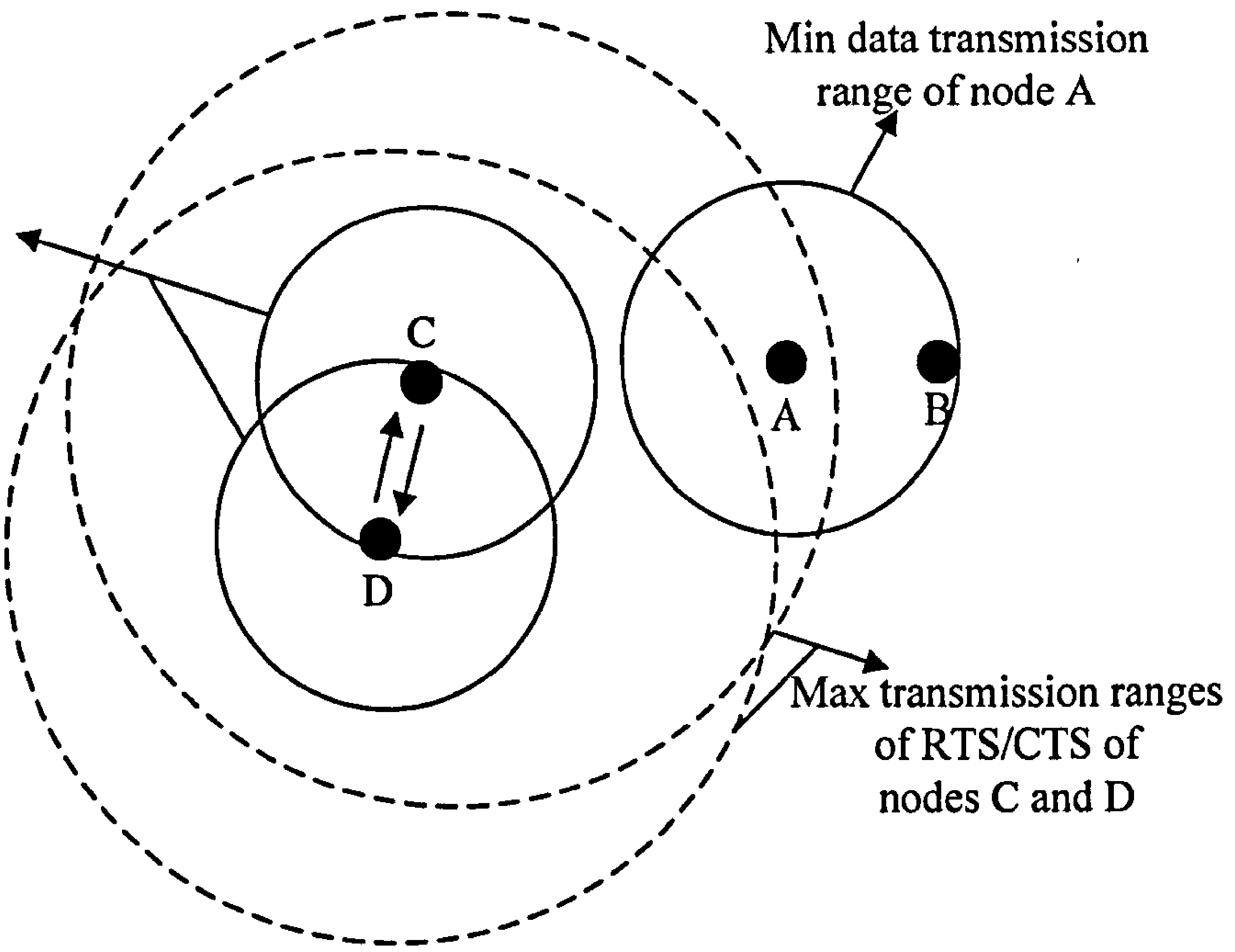


Figure 2.5 Simultaneous communication problem between node A & B

The aim of the variable transmission power approach is to reserve only the required transmission floor in order to maximise the probability of being able to support simultaneous transmissions in the neighbourhood. In addition, minimum power also saves battery power consumption. However the above approach creates collisions and prevents simultaneous transmission as discussed above. Such problems are caused by the hidden nodes, asymmetry in the transmission floors and lack of information about the on-going transmissions to the interferers.

2.2.1.4 Interference-aware MAC schemes

Collisions between the neighbouring nodes can be prevented if the source/relay node knows about on-going transmissions in the neighbourhood. This approach has been described by several authors [29-31]. Figure 2.6 shows a scheme for avoiding collisions. If node D is transmitting to node E, and node A wants to initiate a simultaneous transmission, then the transmission power of node A should not cause interference at node E and node F. Node A needs to be aware of active receivers in the neighbouring region. Hence, in this example, node A's transmission power is limited by the maximum power that node E and node F can tolerate without causing errors. Similarly the maximum transmission power at node C is determined by the maximum allowed power signalled by node E and node F.

Each active receiver measures the carrier to interference (C/I) ratio and determines the maximum additional power that can be tolerated without causing degradation to the received signal.

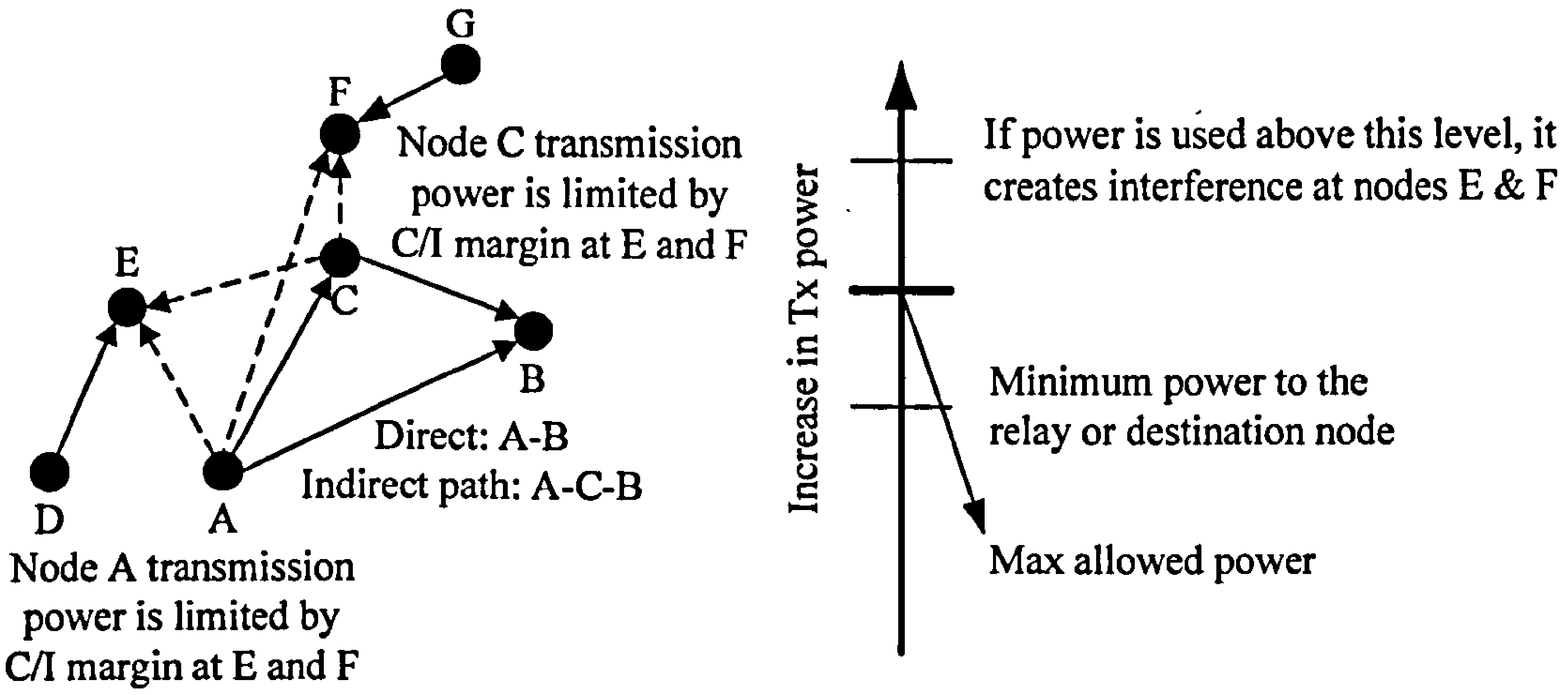


Figure A)
Solid line = data path
Dottled lines = Interference

Figure B)
Power levels of node A

Figure 2.6 Interference aware nodes in the network

Monks et al [29] proposed a Power Controlled Multiple Access (PCMA) scheme in which received power levels from each receiver is periodically broadcasted in the busy tone channel. The busy tone channel is a separate control channel. The source node first monitors the busy tone channel to determine its maximum allowed power by measuring the maximum received power on the busy tone channel. In order to prevent interference with the on-going communication, the minimum power to the relay or destination node should be less than the maximum allowed power.

2.2.2 Routing in ad hoc networks

A second key issue for mobile ad hoc networks is routing. Mobility in ad hoc networks results in changes in the network topology, and this adds complexity to the routing protocol design. For example, referring to figure 2.7a, assume that node A sets up a route via the relay nodes B and C to the destination node D. If the nodes in the ad hoc network are mobile, and that the relay node B moves out of range of transmission from node C, and node C is also out of range of node A, as shown in figure 2.7b. Hence, node A's route to the destination node D is then affected by node B's mobility, and therefore node A needs to initiate a route discovery process.

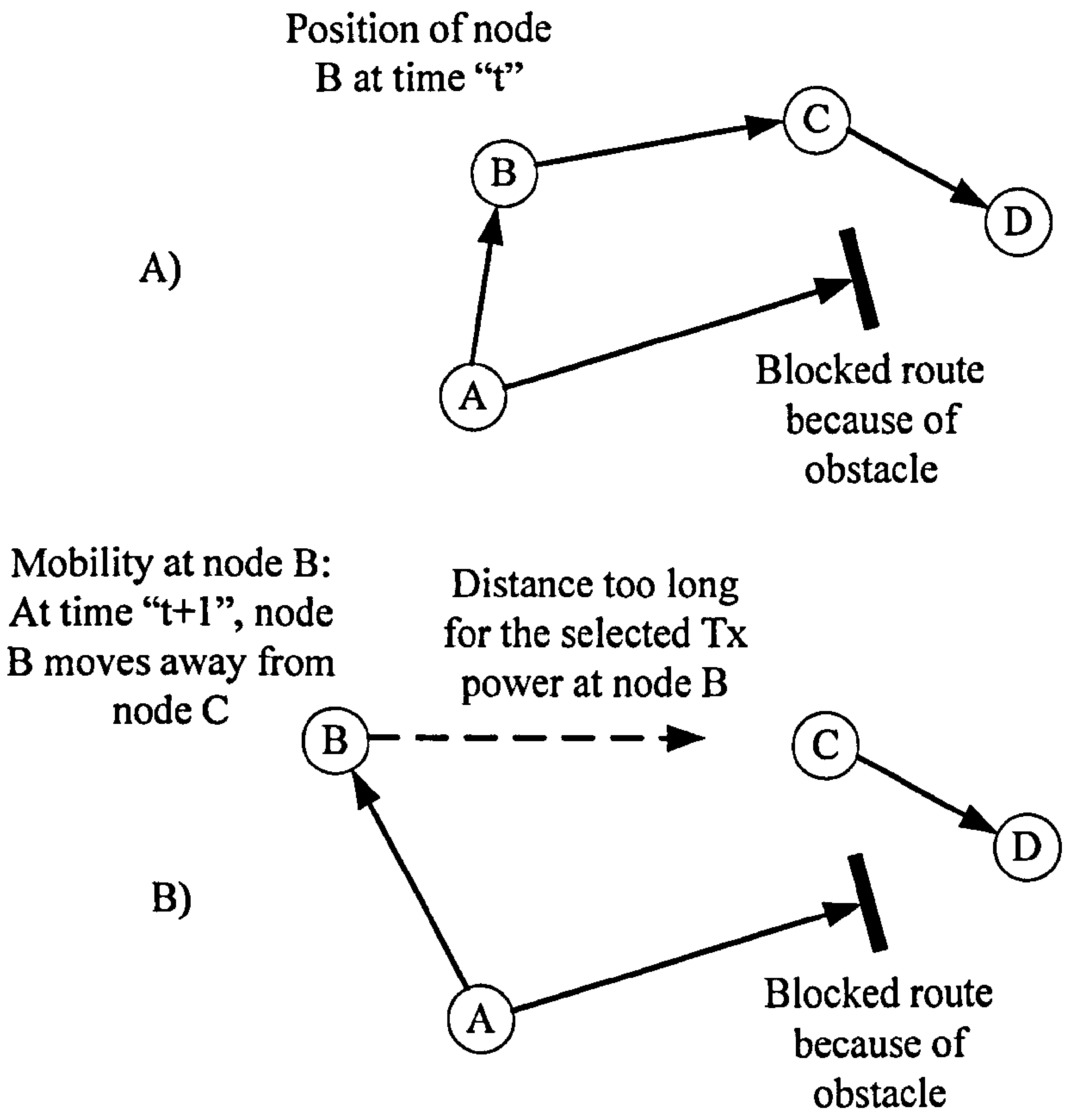


Figure 2.7 Problems in routing data due to change in network topologies

If the terminals in the network are battery powered, then the network topology will also be affected when a node's battery is exhausted because that node is then no longer able to participate in the network. In figure 2.7a), if node C's battery is completely exhausted, then the node C disappears from the network, and node A needs to identify a new route to node D.

The primary objective of an ad hoc routing protocol is to identify the best route to the destination node. In addition, the routing protocol should have minimal control overhead, minimal processing overhead, route maintenance and loop prevention. Routing protocols for wired networks (such as the link state and distance vector protocols) are designed for a static environment where topology changes are infrequent, and the links are highly reliable. However these protocols are not the optimum solution for mobile ad hoc networks due to the continuous changes in the topology [32]. For example, link state routing protocols have a faster route convergence time than the distance vector protocol; however the addition of links or nodes in the link state routing protocol creates more routing overheads which may not be suitable in a mobile ad hoc network.

Various mobile ad hoc network routing protocols have been proposed. They can be divided into two main categories: proactive (table driven) routing protocols and reactive (on demand) routing protocols [33]. Figure 2.8 shows various proactive, reactive and hybrid routing protocols.

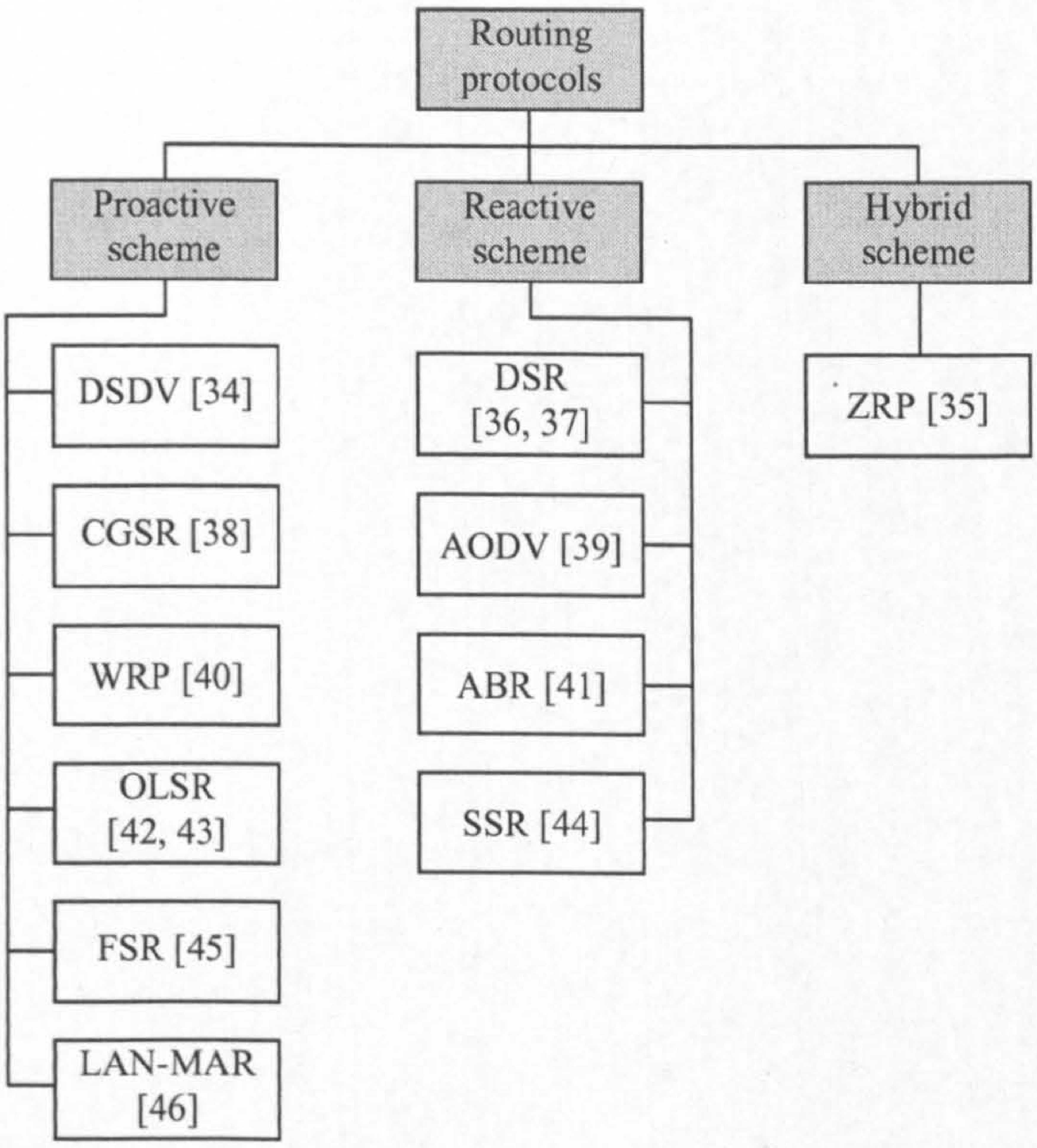


Figure 2.8 Proactive, reactive and hybrid routing protocols

2.2.2.1 Proactive routing protocols

Proactive routing protocols maintain up to date routing information about every node in the network. These protocols are derived from the traditional distance vector and link state protocols. Each node maintains one or more tables to store routing or neighbour information by propagating periodic or triggered routing updates. Destination-Sequenced Distance-Vector (DSDV) [34] is a proactive routing protocol based on the Bellman-Ford algorithm. Every node in the network maintains a routing table of all the possible destinations and the number of hops. Each entry in the routing table is allocated a sequence number which differentiates stale routes from new routes. Each node transmits its routing table periodically in order to keep the routing table up to date and consistent. In order to prevent an excessive amount of routing traffic, full and partial routing updates are broadcast. A full update carries all available routing information, whereas a partial update only carries information which has changed since the last full update. In a network of large number of nodes with mobility, a partial update will broadcast more frequently in the network. In addition, it is assumed that all nodes can receive these broadcasts so that all nodes can maintain up to date routes to all the destination nodes in the network.

The Clusterhead Gateway Switch Routing (CGSR) [38] protocol uses DSDV as the underlying routing scheme however it differs in terms of network organization. In the CGSR scheme, the cluster head controls a group of ad hoc nodes. A distributed cluster head selection algorithm is used for the election of a cluster head node in the group. Two cluster heads are linked through a gateway node. Nodes are considered as gateway nodes if they are within communication range of two or more cluster heads. The source node first sends a packet to its cluster head, then the packet is forwarded to the gateway node, then to another cluster head, until the cluster head of the destination node is reached. Each node maintains two tables; a cluster member table and a routing table. The cluster member table is used to store the destination cluster head for each node in the network. The routing table is used to determine the next hop in order to reach the cluster head. When a packet is received, a node will find the nearest cluster head along the route to the destination in its cluster member table. After that, the node will check its routing table to determine the next hop used to reach the selected cluster head.

The benefit of this approach as compared to DSDV is that the nodes are grouped in to clusters so that channel access and bandwidth utilization can be improved. Spatial re-use can be applied across clusters by allocating different spreading codes (in the case of Code Division Multiple Access) or different frequencies (in the case of Frequency Division Multiple Access). Bandwidth can be conserved by allowing only the cluster head node to broadcast packets within the cluster; however other nodes in the cluster can only forward packets to the cluster head node. Broadcasts from cluster head node exists within the cluster. In the case of inter-cluster communication, the cluster head node forwards packet to the gateway node. The drawback of this scheme is that frequent cluster head changes due to the mobility of the nodes affects the routing protocol performance because nodes are busy in cluster head selection. As a result, nodes will hold packets until the cluster head is selected.

The Wireless Routing Protocol (WRP) [40] is a distance vector routing protocol using the Bellman-Ford algorithm. Each node sends an update message to its neighbours when there is a change in the link. After receiving an update message, neighbours modify their distance table entries and check for new possible routes through other nodes. Any change in the routes is sent back in the update message to inform other neighbours about the change. An acknowledgement is required for each update message. A history of transmitted update messages is stored in the message retransmission list, and this records which update message needs to be retransmitted and which neighbours should acknowledge the retransmission. In the presence of mobility, links between nodes can change frequently which generates update messages to the neighbours. Neighbouring nodes calculates new routes to all the destination nodes and send acknowledgments for each update message. This update/acknowledgement transmission can be continuous depending on the mobility, which increases the routing overhead.

Optimized Link State Routing (OLSR) [42, 43], is a link state proactive routing protocol which exchanges regular topology information with other nodes in the network. Each node computes its multipoint relays from its set of neighbours. The multipoint relays are responsible for performing flooding of control messages in the network, as opposed to a pure link state protocol in which each node floods all the links in the entire network.

Fisheye State Routing (FSR) [45] is a link state proactive routing protocol that maintains topology information at each node and propagates link state updates. FSR uses an efficient way to disseminate routing information by exchanging link state information with immediate neighbours (hop=1) periodically, as opposed to conventional link state where updates are flooded throughout the entire network. The reduction in routing update overhead is obtained by exchanging routing updates with hop ≥2 at a lower repetition rate. Information about the route becomes progressively more accurate as the packet gets closer to the destination. For example, if there is a link changed because a distant node has moved, then this will be notified through periodic routing updates by immediate neighbours. Hence, information can be propagated through the network. FSR uses periodic routing updates which could generate unnecessary routing traffic if links are stable and do not change often.

An enhanced Landmark Routing (LAN-MAR) [46] combines FSR [45] and Landmark routing [47] which is designed for an ad hoc network that exhibits group mobility. Each group has one dynamically elected node serving as a landmark. A distance vector routing protocol such as DSDV [34] is used to propagate routing information between all the landmarks in the network. Local nodes in the group use FSR to exchange routing information regularly. When a node needs to relay a packet to a node within its group, it uses FSR. If a packet is destined for a node outside of its group, it is forwarded towards the destination's landmark. When a packet arrives at the destination's landmark, it is routed using local FSR tables. LAN-MAR reduces routing table size and routing overhead by summarizing routing information for remote groups of nodes. The disadvantage of LAN-MAR routing is that the elected landmark node may lose its connections with other nodes in the group if it moves location. During this time, local nodes in the group will start the landmark node election process. In the transient period, the local node will hold a data packet (if the packet is destined for another group) until the landmark node is elected for the group. Work is in progress to address this issue.

2.2.2.2 Reactive routing protocols

Reactive protocols uses a different approach. Instead of saving all the routes to the destination nodes locally, a route is identified only when there is data to send. When a source node requires a route to the destination, it initiates a route discovery process and the process is completed once a route is formed or all the possible nodes in the network are examined. Dynamic Source Routing (DSR) [36, 37] uses source routing in which the source node determines the complete hop by hop route to the destination node. When data arrives at the source node to deliver to the destination, it performs a route discovery process if the route does not exist in the cache. Route discovery works by flooding route request (RREQ) messages in the network. Route replies (RREP) are generated by the destination node or a node which has a route to the destination. The source node can save several different routes to the destination in the cache. If any link on a source route is broken, this is notified by a route error packet (RERR). The source node removes all the affected links from the cache. Hence, a new route discovery is needed by the source node. The benefit of DSR is that during route discovery process, node learns about other nodes on the route to the destination node. RREQ messages are broadcasted in the network, which means that all the nodes will broadcast this message if the route to the destination node is not available. This creates extra routing traffic in the network.

Ad-hoc on-Demand Distance Vector (AODV) [39] discovers routes on an "as needed" basis through a similar route discovery process to the DSR protocol. However, AODV uses an optimized RREQ flooding scheme called "expanding ring search" in which a time to live (TTL) field in the IP header increases in every RREQ packet until the destination is discovered. The routes are maintained in the routing table with one entry per destination. If a routing table entry is not used recently, then this entry is deleted. In addition, AODV includes destination sequence number. If there are two routes to the same destination, then the fresher route (based on the destination sequence number) is always chosen. During the route discovery, the nodes can learn about routes only to the source of any incoming routing packets being forwarded. As a result, AODV relies more on route discovery which may generate significant routing overhead. AODV does not save alternative routes (if they exist) to the destination node; and if one route does not work, then route discovery is required.

Associativity-Based Routing (ABR) [41] is a source-initiated on-demand routing protocol. In ABR, a route is selected based on the degree of association stability of the mobile nodes. Association stability is defined by the connection stability of one node with respect to another node. A high degree of stability indicates a low state of node mobility, and a low degree of stability indicates a high state of node mobility. Each node periodically sends a beacon to broadcast its existence. When a beacon is received by its neighbours, the associativity value of the current node with respect to the beaconing node is incremented. The route discovery phase consists of a broadcast query (BQ) and an await reply (BQ-REPLY) cycle, and the best route is selected by the destination node which sends a REPLY packet back to the source node to inform of the best selected route. Only one route to the destination is stored in the source node. The ABR protocol uses a different mechanism for route maintenance when nodes are mobile. When intermediate node (IN) or destination node moves in the selected route, then instead of initiating BQ-REPLY cycle, IN's or destination node's immediate upstream neighbour will send a local broadcast query to the destination node to find a new route. The local broadcast query is a controlled broadcast as opposed to full broadcast query, which propagates up to the set number of hops from the upstream node.

Signal Stability Routing (SSR) [44] is an on-demand routing protocol which selects routes on the basis of signal strength between nodes and a node's location stability. There are two components in SSR; dynamic routing protocol (DRP) and static routing protocol (SRP). The DRP is responsible for maintaining the signal stability table (SST) and the routing table. The SST records the signal strength of the neighbouring nodes by a periodic beacon frame transmission at the link layer. The SRP processes packets by searching for a route to the destination node in the routing table. If the route does not exist, then a route discovery is initiated which consists of sending a route request. A route request is forwarded only to the next neighbour if strong signals are received through beacon frames. When a route request is received at the destination node, a route reply is generated and transmitted back to the source node. The DRP updates the routing tables of the nodes along the path. If the source node does not receive RREP packet within a certain time period, then the source node changes the 'preference' field in the header of RREQ packet to indicate that weak signals are acceptable. The main idea behind SSR is to relay traffic through longer-lived routes based on signal strength and location stability. However the performance of the protocol depends upon the mobility and terrain characteristics.

2.2.2.3 Hybrid routing protocols

The Zone Routing Protocol (ZRP) [35] is a hybrid routing protocol that uses both proactive and reactive routing strategies. Each node has a predefined zone centred in terms of the number of hops. Proactive routing protocols are used when maintaining routing information between the nodes within the zone. For nodes outside of a zone, a reactive routing approach is used based on RREQ/RREP packets. The proactive Intrazone Routing Protocol (IARP) is used to maintain routing information within the zone. The IARP can use link state or distance vector routing. For nodes outside of the zone, reactive Interzone Routing Protocol (IERP) is used which uses an on-demand approach to discover routes. The benefit of ZRP is that IERP uses controlled broadcasts for RREQ queries. The scheme uses a multicast-based probing service that the authors refer to as controlled broadcasting, RREQ packets are directed towards the border nodes (edge nodes) of the zone rather than broadcasting to all the neighbour nodes. RREQ queries can be further minimised by increasing the size of the zone. However, this would increase the size of IARP routing table. A care is needed so that increasing the size of proactive zone does not affect the performance of IARP.

2.2.2.4 Comparisons of proactive and reactive routing schemes

Proactive and reactive routing protocols are designed for different network conditions, and applications. If the user nodes are moving rapidly so that the network topology is changing or if it is necessary to establish a route quickly or at frequent intervals, then a proactive routing protocol is used. Proactive routing schemes send regular routing updates and maintain routes for destination nodes without a need of communication. The routes to all the destinations are maintained in a routing table and there is no delay in searching for a route when sending data to the destination node. However the disadvantage of these features is that proactive schemes require extra bandwidth and consume energy due to regular routing updates. In addition, the amount of signalling and the size of the routing table increases with the number of nodes in the network. Some proactive routing protocols scale for use in large networks by reducing the routing packet size and update frequency; other protocols restrict the nodes to propagate routing updates in the network.

If the network nodes are static or very slow moving or where the nodes send data infrequently, a reactive protocol is used. The advantage of a reactive routing protocol is the potential reduction in the routing overhead, as it only selects a route when data needs to be sent. The disadvantage of this scheme is that it can take longer to find a route particularly for networks with a large number of nodes. Table 2.1 summarises the conditions best suited to proactive and reactive routing protocols.

	Data frequency	
Speed	Low	High
Low	Reactive	Proactive
High	Proactive	Proactive

Table 2.1 Selection criteria of proactive and reactive schemes

An ad hoc network can be single hop or multi hop, the topology can be fixed (in the case of sensor networks) or highly mobile. Nodes in an ad hoc network may be battery powered with limited energy or powered by a continuous power supply. In all these circumstances, a routing scheme should have the flexibility to adapt to the network topology and meet the requirements of the application. A proactive or a reactive protocol alone will therefore not be ideal for all network situations. A solution to this dilemma is to implement a hybrid strategy in which the routing protocol can function in a proactive or a reactive mode.

2.2.3 Transport control protocol

A third key issue for mobile ad hoc networks is the use of a transport layer protocol. The transport control protocol (TCP) was originally designed for use in wire line networks to provide reliable delivery of data [48, 49]. If TCP is used without any modification in mobile ad hoc networks, then a serious drop in the throughput is observed. There are several reasons for such reduction in throughput.

2.2.3.1 Effect of high Bit Error Rate (BER) and network partition

Variable attenuation and fading in the radio link causes a high bit error ratios. The errors cause data packets to be corrupted which results in lost TCP segments or ACK. When the TCP sender does not receive the expected ACK within the time limit, then TCP interprets this as congestion in the network, and invokes the TCP congestion algorithm (retransmits the segment and reduces congestion control window threshold) [50]. Frequent errors will continue to use TCP congestion algorithm which results in a reduction in data flow.

During data transmission, if the link is disconnected, for example the receiver is out of the range of the transmitter, then the sender's packet will be dropped. If these segments are re-transmitted then these segments will be dropped again because of the link failure. In addition, using TCP, the re-transmission time-out at the sender is doubled with each unsuccessful retransmission attempt, and hence several failure retransmissions could make the sender node inactive for a short time, even when the sender and the receiver link is reconnected.

Packet loss due to congestion or link failure can be differentiated with the help of Explicit Congestion Notification (ECN) mechanism [51] and Explicit Loss Notification (ELN) mechanism [52, 53] respectively. In a multi hop route, if a packet is lost due to limited buffer capacity at an intermediate node, then this is notified to the TCP sender to reduce the window size. If a packet is lost due to link failure, the receiver informs the sender TCP to retransmit lost segment having the same window size.

2.2.3.2 Effect of route recomputations

The routing protocol at the network layer finds a new route if an old selected route is no longer available. Finding a new route will take some time (if an on-demand routing protocol is used). There is a possibility that the time it takes to discover a new route will take longer than the TCP retransmit timer, and hence TCP will invoke the congestion control algorithm.

TCP-F [54] and ATCP [55] schemes have been proposed in which intermediate nodes notifies the sender node about any route failure or delay due to the searching of a new route. The TCP at the sender side goes into snooze state, stops sending further segments and freezes all of its timers.

2.2.3.3 Effect of MAC protocol on TCP

TCP problem in a multi hop ad hoc network using standard IEEE 802.11 MAC protocol has been reported by several authors [56-58]. The problem is mainly due to the exposed node problem. This is illustrated in figure 2.9. Suppose, node 5 sends data packets to node 3 via node 4 at time "t". Each node reserves the medium with the help of RTS/CTS frame exchanges. Assume, each node can hear intermediate neighbour node transmissions only; so node 5 transmission can be heard by node 4 only. Suppose node 5 sends another data packet to node 3 at time "t+1". It initiates RTS/CTS frame exchange with node 4. At the same time (t+1), node 2 initiates data transmission with node 1 and sends a RTS frame to node 1. Node 4 and node 1 will reply CTS frames to node 5 and node 2 respectively. However, data transmission between node 2 and node 1 leaves node 3 as an exposed node. As a result, when node 4 sends RTS to node 3, RTS packet will interfere with node 2's transmission. If data transmission between node 2 and node 1 goes for a longer period, then RTS transmission from node 4 continues to interfere at node 3. When the RTS frame fails seven times then the MAC layer reports a link breakage.

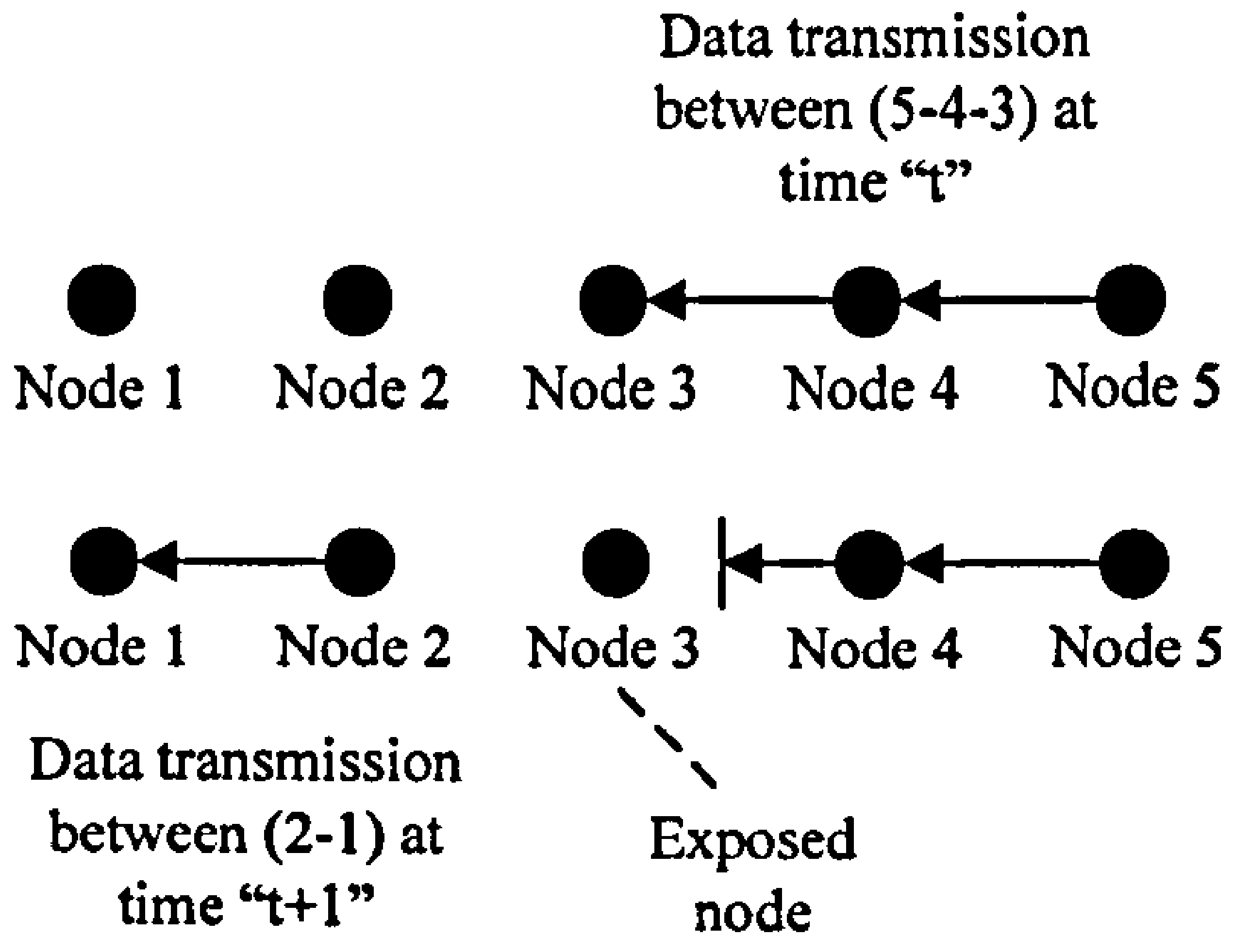


Figure 2.9 TCP problem in multi hop ad hoc network

The original TCP protocol requires ACK of the transmitted TCP segment. In the IEEE 802.11 network, each node reserves data channel through RTS/CTS. However, when a receiver transmits ACK for every TCP segment, then RTS/CTS control frames need to be exchanged before transmitting each TCP ACK which generates additional overhead at the MAC layer. Continuous reservation of the channel for the transmission of TCP ACK will prevent data transmission for other nodes in the neighbourhood.

Channel reservation time can be minimised if a receiver merges several acknowledgements into a single TCP ACK. Request for comments (RFC 2581) [59] recommends that a receiver should combine two ACKs and the maximum delay to transmit ACK should not exceed 100ms. Oliveira et al [60] explains that a fixed timeout at the TCP receiver will generate retransmissions at the TCP sender because the packet inter-arrival time changes with the channel data rate and also with the amount of traffic going through the relay nodes (assuming a finite buffer capacity at the relay nodes) and the number of relay nodes in the route. The authors proposed a TCP-DAA (Dynamic Adaptive Acknowledgement) scheme which enhances RFC 2581. This scheme combines more than two ACKs and has a dynamic timeout interval. Dynamic timeout at the TCP receiver is calculated on the basis of incoming packet inter-arrival time. The delays at the receiver should be enough to avoid retransmissions by the TCP sender.

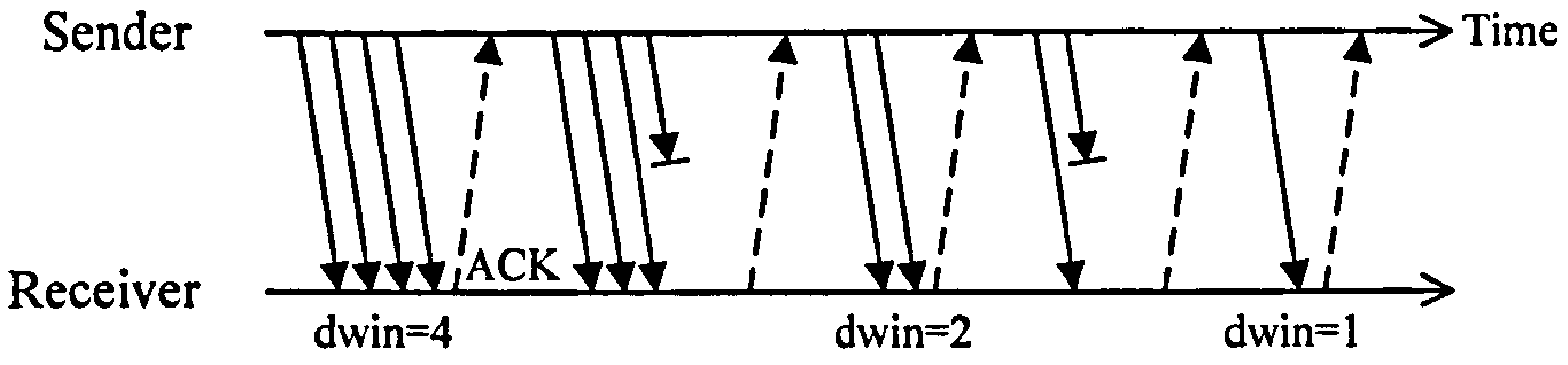


Figure 2.10 TCP-DAA approach

Figure 2.10 shows the behaviour of TCP-DAA approach. In the presence of no losses, the receiver merges four ACKs into a single ACK. In addition, the TCP receiver also calculates the timeout interval based on the inter-arrival time of the packet. If the segment is lost in transit, then the receiver sends an ACK immediately before its timer expires and reduces delaying window to 2. In case segments continue to be dropped, the receiver reduces the delaying window to 1 in order to avoid timeout at the receiver.

In a multi hop ad hoc network, a packet could encounter variable delay due to the number of hops. Moreover, a variation in the number of hops (due to a change of route) during data transmission could affect delay. The receiver in the TCP-DAA approach can adjust the ACK timeout; however a feedback mechanism is needed so that the sender is aware of delays at the receiver so that re-transmissions can be avoided.

TCP schemes discussed above can be used in an ad hoc network. For example, if nodes in a network are experiencing congestion, the ECN scheme can be used to slow down sending rate. If a packet is lost due to a route failure, then ELN scheme would help the TCP sender node to retransmit lost segment instead of invoking congestion control algorithm. If a route failure or change of route happens frequently, then schemes such as TCP-F and ATCP can be used to freeze TCP timers to avoid unnecessary retransmissions.

Exposed node problem in CSMA/CA with RTS/CTS affects TCP operation in a multi hop ad hoc network. One of the possible solutions is to use interference aware MAC protocol, however work is going on to identify medium access protocol for TCP in a multi hop ad hoc network.

2.2.4 Energy conservation

Most of the nodes in the ad hoc network are powered by battery with limited battery charge. Each node is responsible for performing the necessary routing and signalling. Node failure due to a lack of available battery charge leads the network partitioning which would isolate groups of nodes. Figure 2.11(a) shows the connectivity of nodes in the network. It is assumed that nodes are using maximum transmission power. Figure 2.11(b) shows that if the batteries of some nodes are over utilized, then there is a possibility that network could be divided into sub-network. Hence minimising the battery consumption is very important for ad hoc networks.

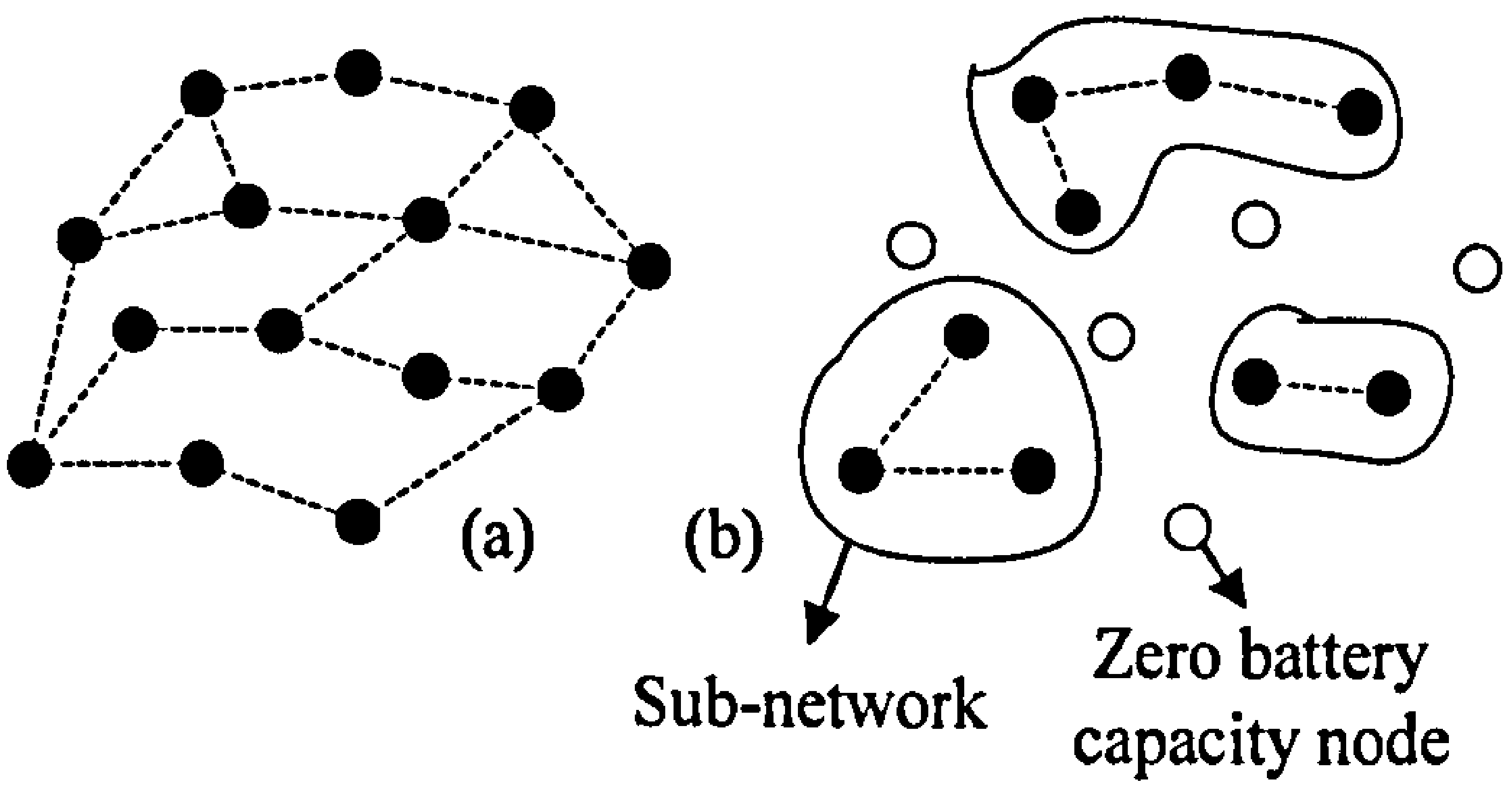


Figure 2.11 Network partitioning due to node failure

Protocols at all layers can influence energy consumption. Reliable communication between source and destination can save MAC and TCP level re-transmissions. At the network layer, the routing protocol could change routing updates from continuous to on-demand, or select a path based on energy related metrics such as end to end minimum power and the node's available battery. Energy conservation is discussed further in chapter 4.

2.3 Summary

This chapter discusses the key design issues for mobile ad hoc networks. At the medium access layer, RTS/CTS can be used to overcome hidden node problem experienced with CSMA/CA. Transmitting RTS/CTS with maximum power maximises the range of communication, but it also maximises the area over which other nodes are prevented from transmitting. Transmission power needs to be managed to maximise connectivity but to minimise interference so that nodes can support simultaneous transmissions. Varying transmission power of control signal and data packet using CSMA/CA with RTS/CTS creates collision and prevents other nodes to communicate. An interference aware MAC protocol such as busy tone scheme can be used to find the optimum power levels to minimise interference.

A range of proactive and reactive routing protocols for ad hoc networks have been described. The objective is to be able to find a route with minimum delay and minimum signalling overhead. Reactive protocols are best suited to networks where the nodes are slow moving with low traffic frequency. Proactive protocols are better suited for networks with faster moving nodes or when there is regular traffic from all the nodes. Networks may often have a combination of these conditions and therefore a hybrid solution that can operate in either a proactive or reactive mode is proposed as an optimum and generic solution.

Standard TCP does not work well in a multi hop ad hoc networks as packet loss due to signal level variation can be interpreted as congestion and lead to re-transmission and further packet loss. ECN and ELN can be included to distinguish congestion and packet loss. TCP-F and ATCP can be used when there is a delay in handover to another route in order to freeze TCP timers at the sender side to avoid re-transmissions. TCP with an interference aware MAC protocol can be used to avoid exposed node problem.

The transmission power, the route selection and the amount of re-transmission due to packet loss all influence the energy consumption and hence the lifetime of an ad hoc networks. This should be taken into account in selecting the protocols and mode of operation at each layer.

3. Cross Layer Design

3.1 Cross layer dependencies in wireless systems

Fixed wireline data networks generally adopt a layered approach to support the functions required for data transmission due to better inter-networking, standard interfaces and modularity. The aim is to group associated functions into a layer; each layer provides services to the layer above but the higher layer does not need to know how the services are implemented and layers can therefore operate independently. Communication takes places between the same layer on different hosts and protocols for this communication have been standardised. The advantage of using standard protocols is that equipment suppliers can design their equipment in the knowledge that their equipment will be able to communicate with other equipment designed using the same layered approach. The TCP/IP reference model [16] defines the most common layered architecture and this has proved to be highly successful for internetworking.

Over the last decade there has been a tremendous growth in wireless data networks and there has been considerable discussion about the suitability of the layered architecture and the TCP/IP protocols for these wireless networks because the transmission characteristics vary with time. The following sections outline the problems that may arise when the TCP/IP layered approach is used in wireless networks.

3.1.1 Issues with wireless network using the layered approach

- **TCP throughput problem**

As discussed in section 2.2.3 that the TCP was originally design for wired networks. In a wireless network, the links are subject to variable attenuation and fading which can also lead to packet loss. TCP interprets a packet loss as an indication of congestion and invokes a congestion avoidance algorithm. It is therefore not possible to distinguish between packet loss due to congestion and packet loss due to a reduction in the received signal strength, and TCP would invoke congestion avoidance in both cases even though it is not an appropriate response to attenuation and fading.

- **Medium access control**

A wired network uses fixed power to transmit data. If a fixed transmission power is used in a wireless network, then the transmission footprint may be larger than required and this can cause interference or unnecessarily prevent access to other nodes. In addition, the power consumption at the transmitter may be higher than necessary which is particularly important for battery operated nodes. If the transmit power is variable, then the transmission footprint can be matched to the required range and power consumption can be minimised. Interference aware MAC protocols such as busy tone are successful for simultaneous communication (see section 2.2.1).

- **Mobility**

Mobility in wireless networks affects the performance of various layers in the protocol stack. For example, at the physical layer, mobility can cause the signal to noise ratio to vary with time; at the MAC layer, mobility affects which nodes are accessible to other nodes at a given point in time; at the network layer, mobility affects the connectivity between the nodes (topology); at the transport layer, mobility can lead to segment loss (which can be misinterpreted as congestion).

The above discussion highlights the problems that can arise when a layered approach is used in wireless networks. The impact of a time varying channel can influence all the layers. Hence optimum performance cannot be achieved if each layer operates independently. An interchange of information between layers is needed to optimise the system performance. This is referred to as cross layer design [18, 61, 62].

3.2 Potential cross layer solutions

Several applications of cross layer design to wireless networks have been reported in the literature and these are outlined in the following sections.

3.2.1 Explicit Congestion Notification (ECN)

The TCP with Explicit Congestion Notification Scheme (TCP-ECN) was originally developed for wired networks in order to notify the TCP layer at the source node about the congestion. The TCP layer at the source node detects dropped packets either from the receipt of three duplicate ACK packets or after the timeout of a retransmit timer, and responds by reducing the size of the congestion window [50]. Packet drops are mainly due to the queue overflow in the router. Hence packet drops and packet retransmissions can result in noticeable delays which can be a problem for delay-sensitive applications.

To address this problem, Floyd [63] and Kunniyur et al [51] discussed the use of the ECN mechanism in the TCP/IP protocol. ECN messages are generated by an IP router that carries TCP/IP traffic. Each router monitors the average queue size and when reaches its pre-defined limit, then sets the ECN field in the outgoing packet header. The addition of ECN to the IP packet header is explained in RFC3168 [64]. Bits 6 and 7 in the IP version 4 (IPv4) Type of Service (TOS) octet (defined in RFC791 [65]) are designated as the ECN field. However, ECN fields are further discussed in RFC2474 [66] and RFC2780 [67]. RFC3168 proposed the use of ECN in TCP. In this proposal, bit 9 in the Reserved field of the TCP header (defined in RFC793 [48]) is designated as the ECN-Echo flag. In essence, the ECN bit is used for signalling between routers and receiver, and the ECN-Echo bit is used for signalling between sender and receiver. Figure 3.1 explains the TCP-ECN mechanism; when the TCP layer at the receiver receives a data packet with the ECN bit set in the packet header, the receiver sets the ECN-Echo bit in the next outgoing TCP ACK packet to the sender.

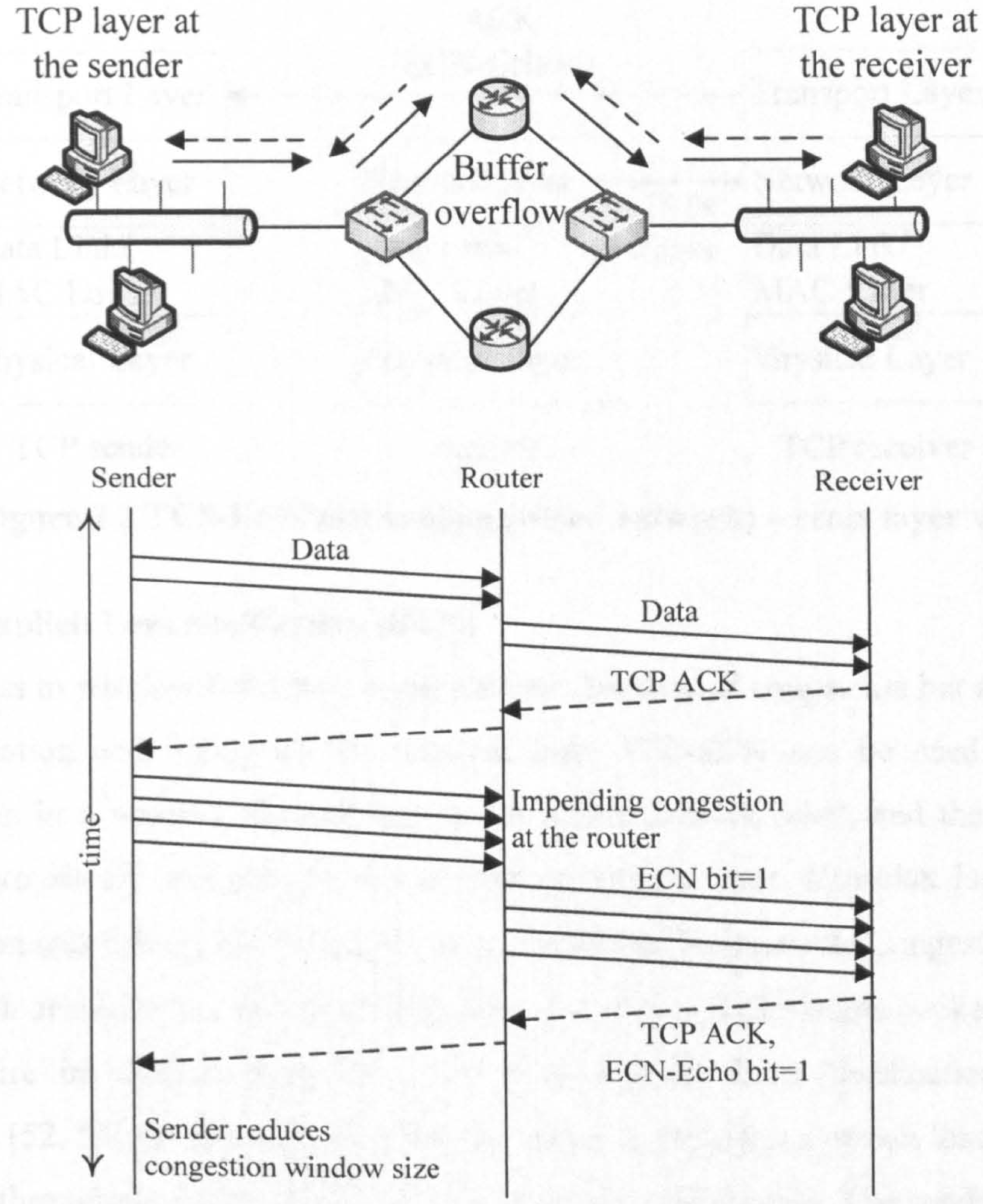


Figure 3.1 TCP-ECN mechanism in the TCP/IP network

Figure 3.2 shows the TCP-ECN mechanism in terms of cross layer view of the TCP/IP layered model. A router detects incipient congestion due to buffer overflow, and sets the ECN field in the IP header at the network layer. Such information is passed to the network layer of the receiver. The network layer at the receiver informs the transport layer about the ECN bit, and hence ECN-Echo bit at the transport layer is set in the next TCP ACK to the sender. The sender reduces the congestion window when a TCP ACK with an ECN-Echo bit is received.

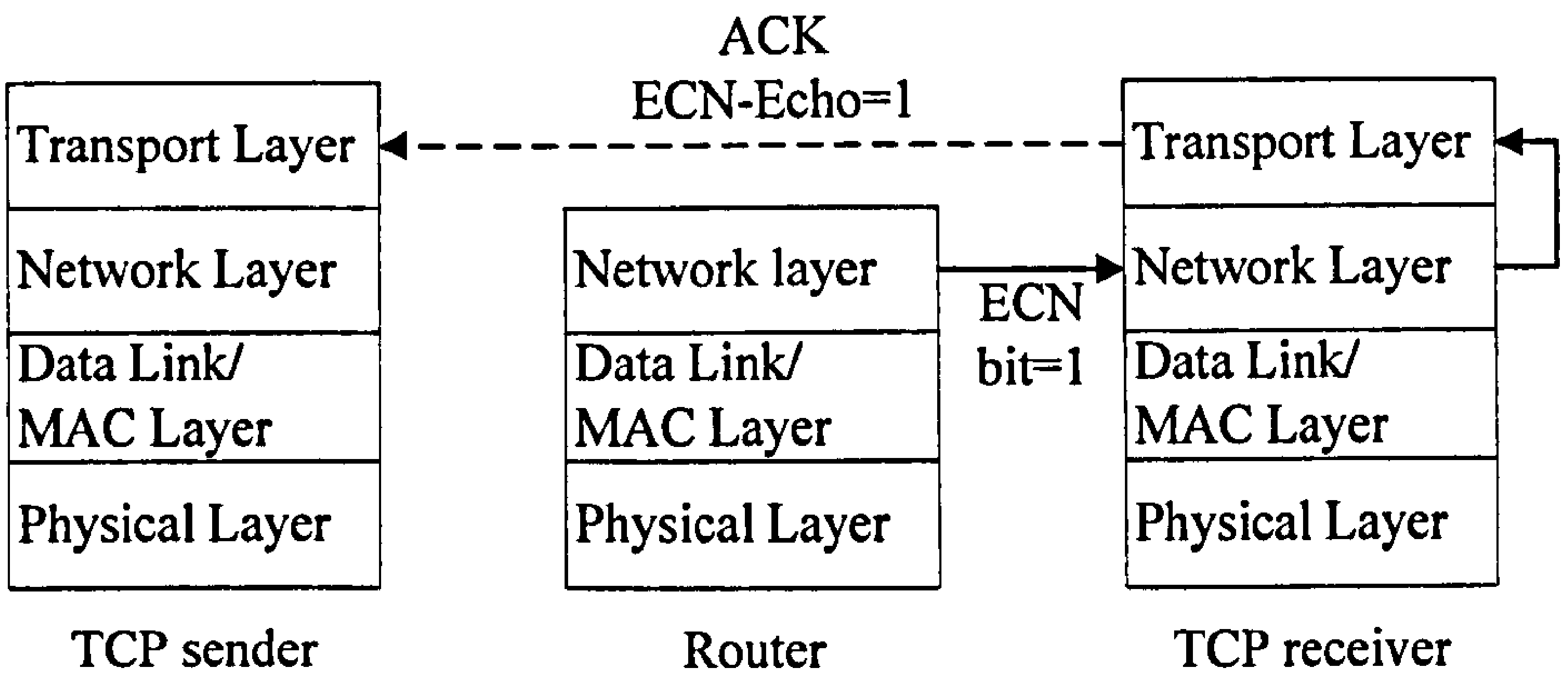


Figure 3.2 TCP-ECN mechanism (wired network) – cross layer view

3.2.2 Explicit Loss Notification (ELN)

Packet loss in wireless links may occur not only because of congestion but also because of attenuation and fading of the wireless link. TCP-ECN can be used to identify congestion in a wireless network (as in the wired network case), and the sender can react appropriately and reduces the congestion window size. If packet loss is due to attenuation and fading, the TCP layer at the sender still reduces the congestion window size which decreases the network throughput. In order to differentiate packet loss due to link failure in wireless networks, TCP with Explicit Loss Notification (ELN) is proposed [52, 53]. In this scheme, when the receiver identifies a packet loss, it sends an ELN bit through the TCP ACK packet back to the sender node. The sender node then retransmits the lost packet rather than reducing the congestion window size. This process is explained in figure 3.3; when a segment is lost, the receiver sets the ELN bit in the TCP ACK and the sequence number of the lost segments, and sends this information back to the sender. In response, the sender retransmits lost segments and avoids unnecessary window reduction. Hence, packet loss due to congestion and link failure can be differentiated through the ECN and ELN mechanisms respectively.

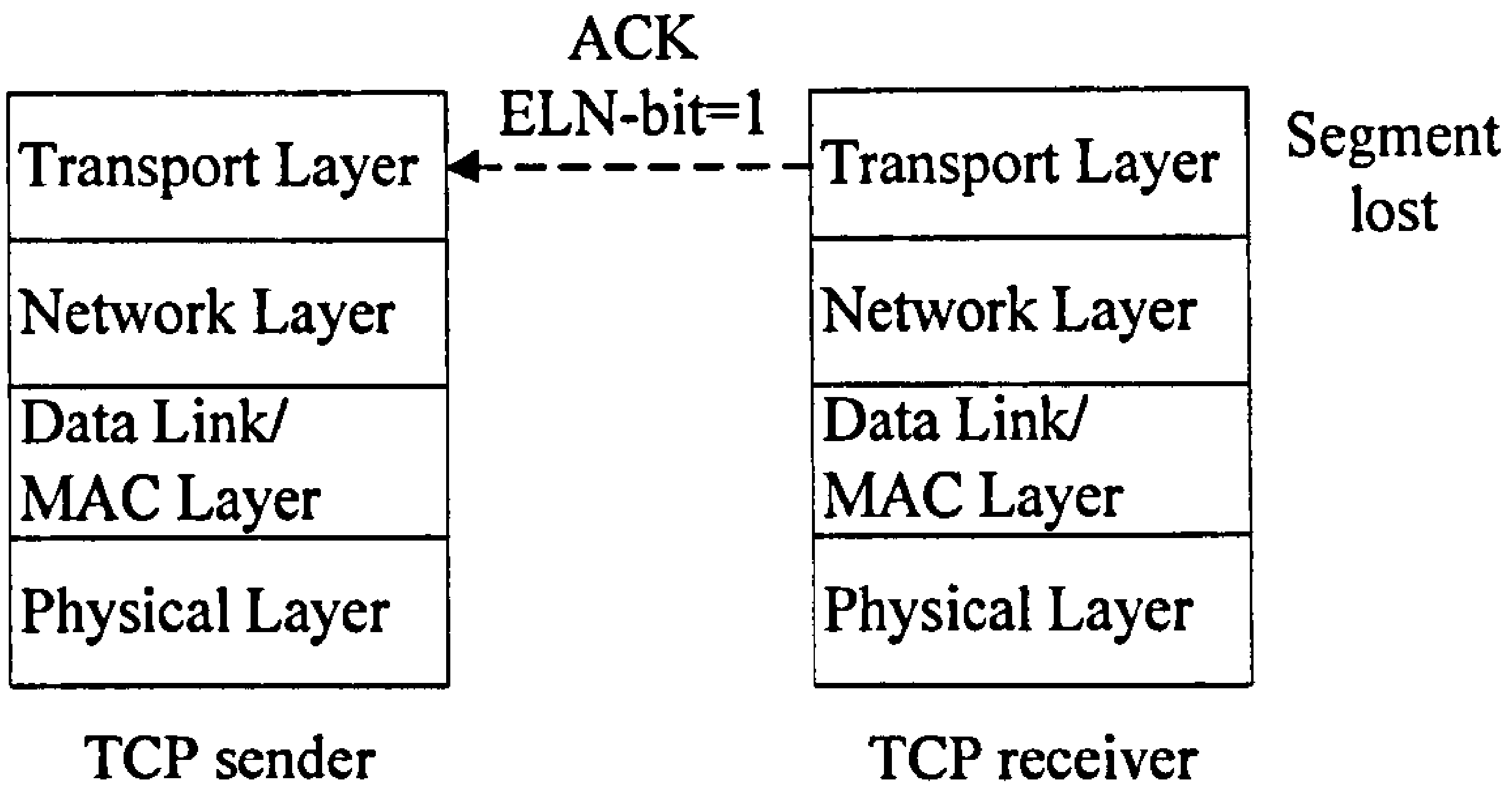


Figure 3.3 TCP-ELN mechanism (wireless network) – cross layer view

3.2.3 TCP for ad hoc networks

As discussed in sections 3.2.1 and 3.2.2, TCP with ECN and ELN enhancements can be used in wireless networks to separately identify packet loss due to congestion and fluctuations in received signal strength (RSS). Such schemes are used in cellular networks where destinations are 1-hop away (mobile station – base station). There are additional issues using TCP in ad hoc networks for routes that use more than 1-hop (multiple relays). Each relay can be mobile which can cause frequent link breakages; moreover, the delay in discovering new routes can cause the expiration of the TCP timer.

Chandran et al [54] proposes a feedback scheme (TCP-F) to improve the TCP performance in a multi hop ad hoc network, which is based on the interaction between the network and the transport layer. Figure 3.4 illustrates this scheme. When an intermediate node detects route failure, the network layer explicitly sends a Route Failure Notification (RFN) packet to the transport layer of the source node. On receiving the RFN, the source node goes into a snooze state and stops sending further segments and freezes all of its timers. When the node which previously forwarded an RFN to the source node discovers a new route to the destination node, it sends a Route-Re-establishment Notification (RRN) packet to the transport layer of the source node to change it from snooze to active state and resume the transmission.

A similar feedback approach (ATCP) is proposed in [55], in which link failure notifications from the intermediate nodes are notified by an Internet Control Messaging Protocol (ICMP) "Destination Unreachable" message. On receipt of this message, the

TCP state at the sender is frozen until a new route is found. In addition, an ECN mechanism is used [63] to identify congestion in the network.

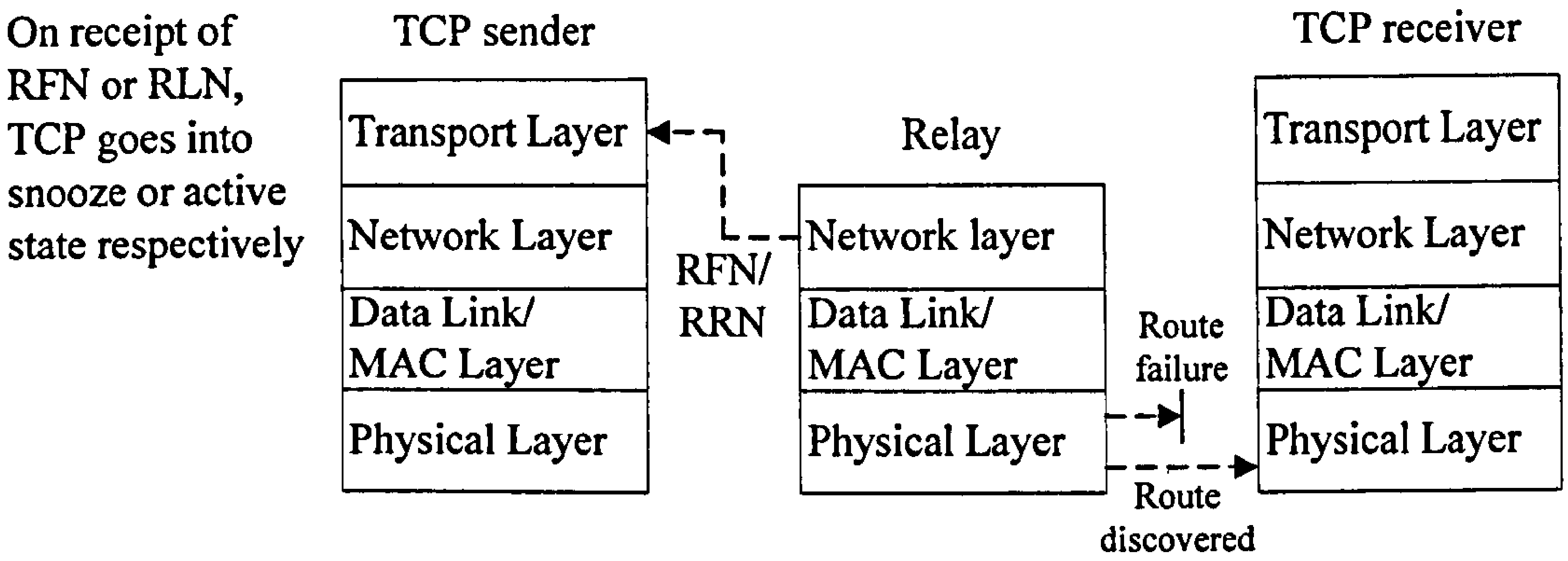


Figure 3.4 Feedback from network to transport layer in the TCP-F scheme

3.2.4 Adaptive modulation

The received signal to noise ratio in a wireless link can fluctuate with time due to variable attenuation and multi path fading. This results in a corresponding fluctuation in the link bit error rate. Adaptive modulation can be used to combat this problem [68-75]. Figure 3.5 shows the adaptive communication system model in which the signal strength at the receiver is sent back to the transmitter through the return path and the transmitter then selects an appropriate modulation to maintain the received signal to noise ratio. This avoids a packet loss due to a low RSS and maximises the network throughput. Adaptive modulation is used in IEEE 802.11 [12], in which the sender keeps track of the RSS measured from the beacon frames sent by the Access Point (AP) at a fixed transmission power.

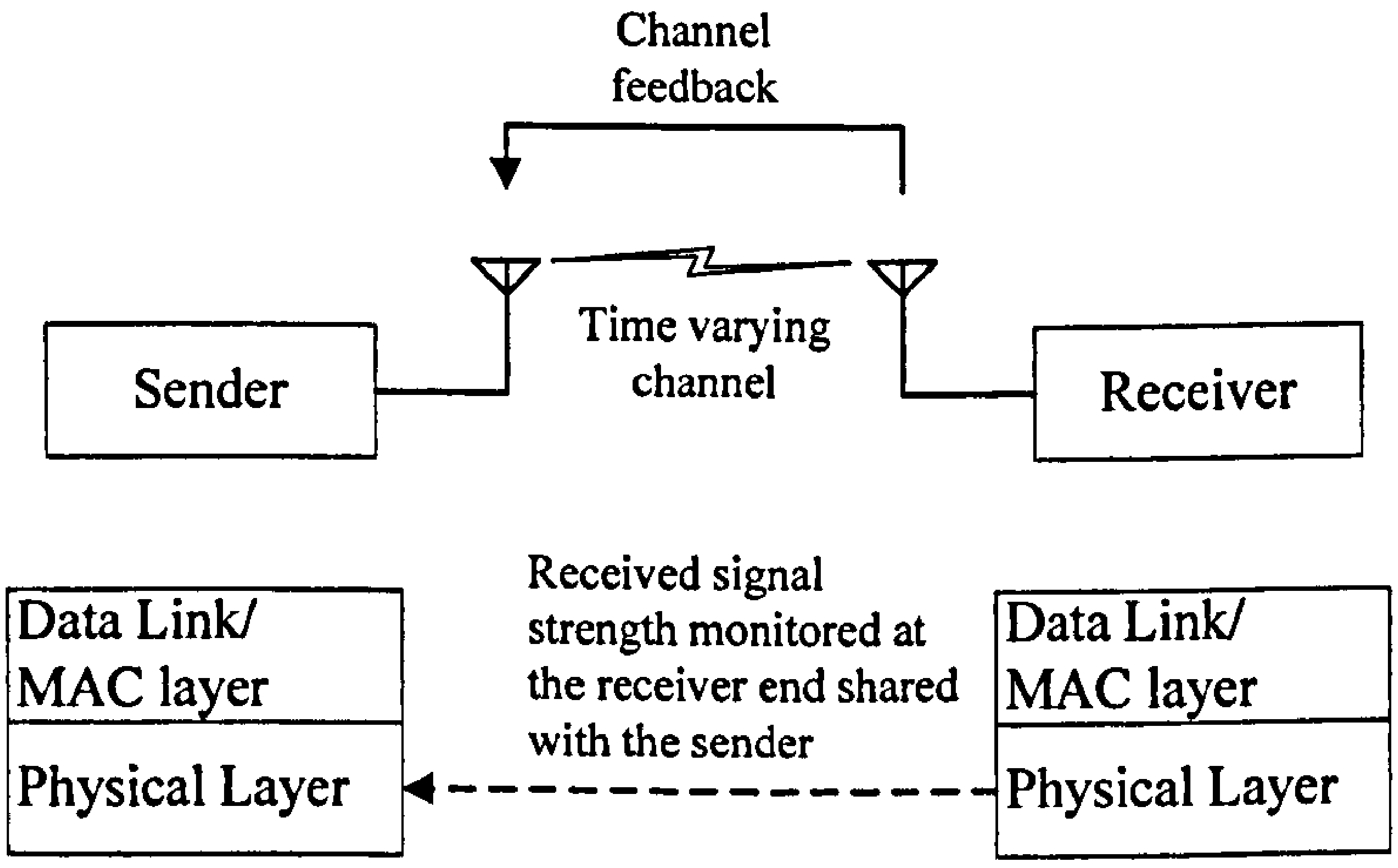


Figure 3.5 Adaptive modulation – cross layer view

Another approach is reported in [76] in which the physical layer at the receiver sends the signal to noise ratio (SNR) to the application layer at the source node. The application layer then informs the PHY layer to selects the optimum modulation and to the MAC layer to select the MAC frame size based on the required application throughput. This is illustrated in figure 3.6.

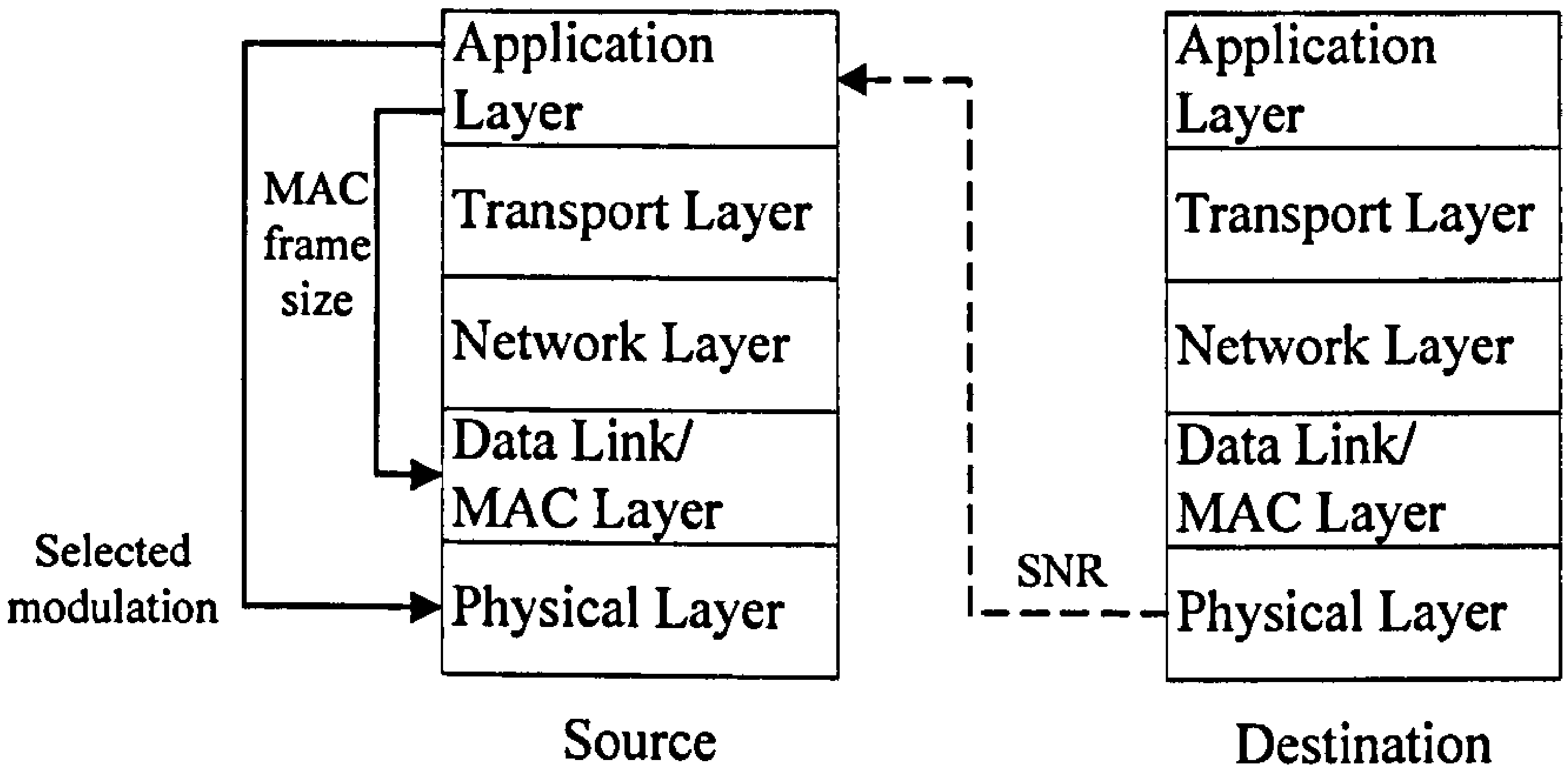


Figure 3.6 Information sharing between physical and application layers

3.2.5 Channel aware access

In a cellular network with a fixed base station and a number of mobile users, the data packets from the Internet to the mobile station are first queued at the base station. The base station then schedules these data packets to the mobile stations. The scheduling can be done by allocating frequency and time slots to each mobile in the round-robin fashion, where users are periodically allocated slots irrespective of whether there is data to be sent to the particular mobile station or not. This leads to inefficient bandwidth utilization due to idle mobile stations. This problem can be solved by transmitting channel state information in the uplink packets from mobile to the base station. This technique is called multi-user diversity [17, 77] and is implemented in the High Data Rate (HDR) versions of Code Division Multiple Access (CDMA2000) [78] and Enhanced General Packet Radio Service (EGPRS) in the Enhanced Data Rates for Groupe Spécial Mobile Evolution (EDGE) [79]. With the knowledge of channel state information and the demand of each user at the mobile station, the base station allocates extra time slots to a mobile station which experiences a good channel state. Hence, downlink and uplink transmissions can take advantage of channel state knowledge at the physical layer of the mobile stations and can share this with the link layer of the base station for efficient scheduling. In the presence of a low signal to noise ratio at the mobile station, both downlink and uplink transmissions can be postponed in order to prevent unreliable data transmission and interference to other mobile stations.

In an ad hoc network, the interaction between the network and the MAC layer is fundamental to the performance of MAC protocol. Higher transmission power increases the connectivity and larger number of potential next hop neighbours, but at the MAC layer, this reserves a larger floor area, and prevents simultaneous communication in the neighbourhood, and hence decreases the network throughput. By the same reasoning, reducing the transmission power may increase the network throughput but at the cost of the limited number of next hop neighbours.

Muqattash et al [80, 81] proposed a Power Controlled Dual Channel (PCDC) medium access protocol in which transmission power levels are adjusted without affecting network connectivity. In this scheme, two channels are used; data and control channels. Each node listens for frames such as RTS/CTS in the control channel which are transmitted at the fixed and known power. On the reception of such frames, the received power is calculated and hence path loss is estimated. Using this value of path loss, the transmission power is calculated which is used to construct the connectivity set.

The connectivity set of each node contains only the neighbouring nodes with which direct communication requires less transmission power than the multi hop communication via any other node that already exists in the connectivity set. In the connectivity set, the transmission power required to communicate with the highest path loss node is used to transmit data on the data channel.

Ad hoc routing protocols such as DSR and AODV use RREQ/RREP packets to discover routes to the destination node. Such RREQ packets are flooded in the network with the fixed power. However, in the PCDC scheme, RREQ/RREP packets are transmitted at the power which is only flooded in the connectivity set, and hence limits the number of broadcasts and reduces the power consumption. Interlayer communication in the PCDC scheme between the physical, MAC and network layers is illustrated in figure 3.7. When RREQ/RREP packets are generated at the network layer, the MAC layer controls the power used to transmit the packets and instructs the physical layer about the required transmission power.

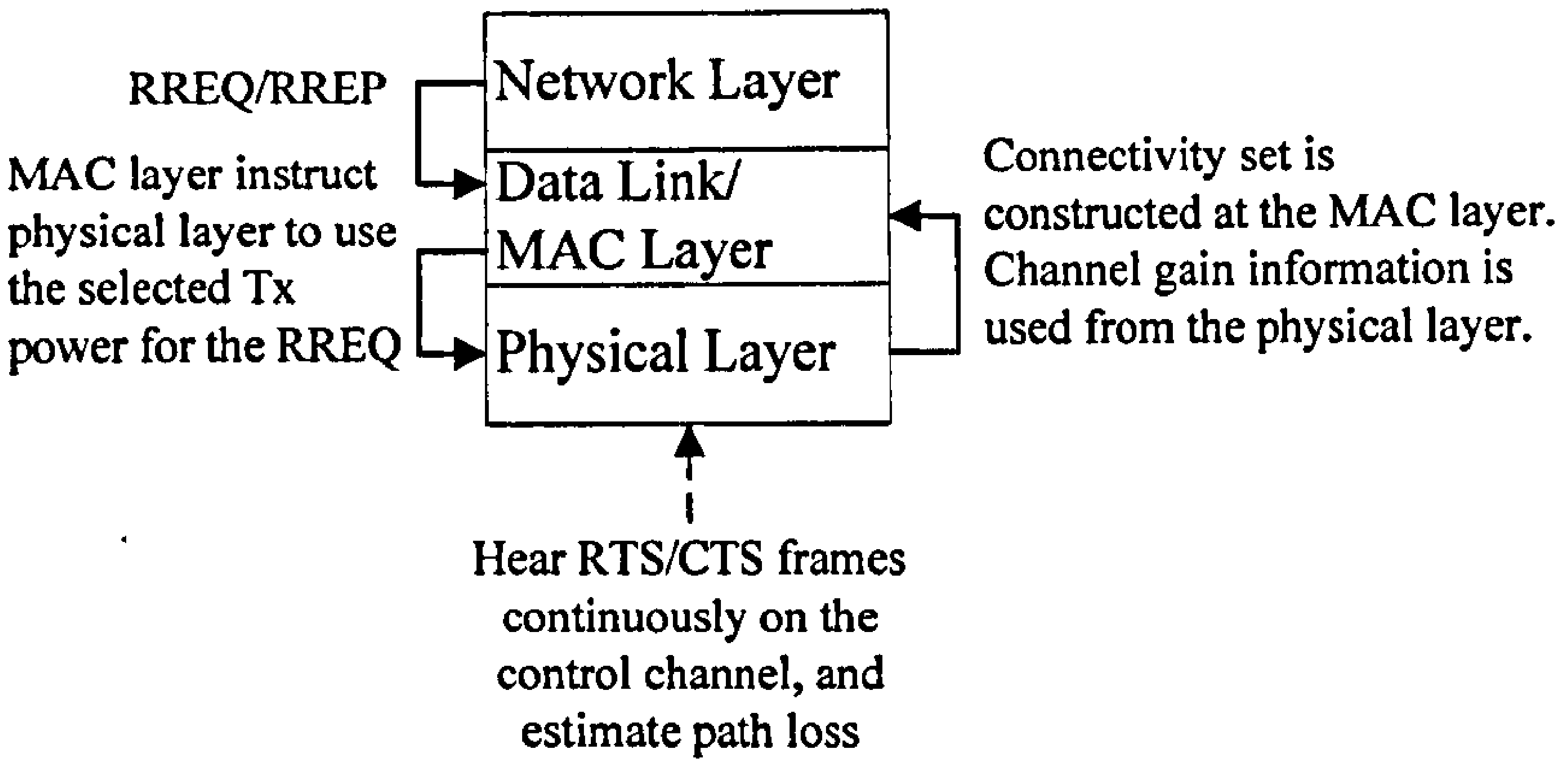


Figure 3.7 Interlayer communication in the PCDC MAC protocol

3.2.6 Link prediction

Qin et al [82] describe a link prediction method in which the physical layer of the receiver monitors the strength of the received packets, and when the strength goes below the threshold, the physical layer sends a message to the network layer to send a prediction route error message to the source node, see figure 3.8. As a result, the source node will find another route from the cache or will broadcast a route discovery packet.

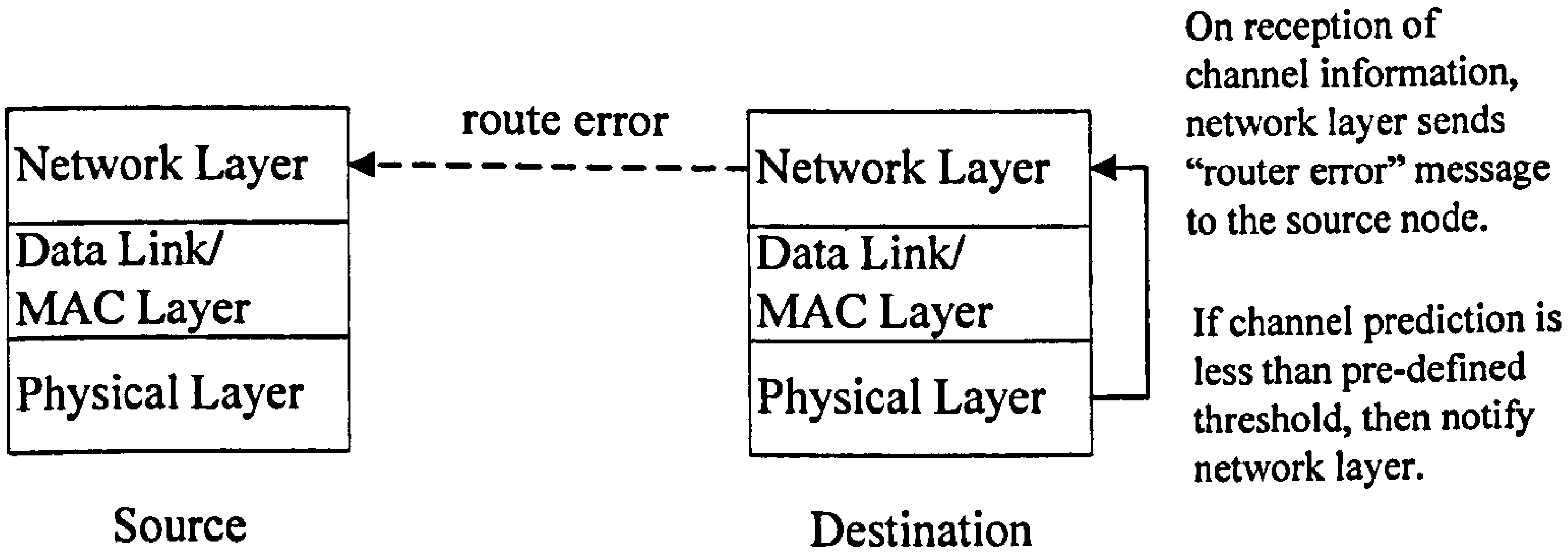


Figure 3.8 Link prediction – cross layering between physical and network layer

Pham et al [83] describe a cross layer scheme in an ad hoc network in which MAC, routing, transport and application layers get the channel status information from the physical layer and halt the transmission when the received signal to noise ratio is low. The receiver station performs the prediction of channel quality which is based on the past received signal strength measurements and determines whether the next packet transmission will be received correctly. If the predicted received signal powers are less than a pre-defined threshold, then the receiver may not successfully receive new packets. The receiver notifies the sender station (by setting a flag in the header of CTS or ACK) to stop the transmission. Upon receiving this information, the physical layer at the sender shares the information with upper layers, and hence the upper layers will halt the transmission, see figure 3.9. Such a scheme will prevent unnecessary transmission which reduces unnecessary power consumption and interference.

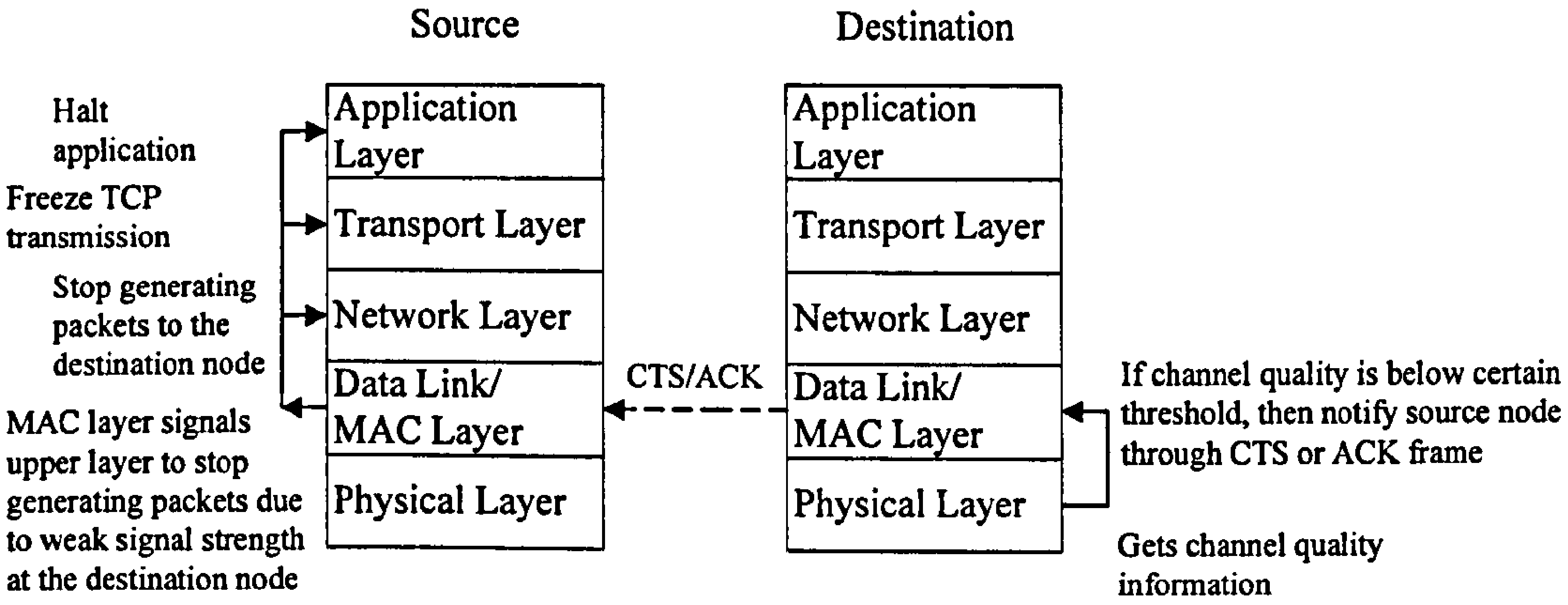


Figure 3.9 Link prediction – cross layering between physical and higher layers

The above mentioned cross layer techniques in the literature can be used all together in an ad hoc network. For example, the ECN mechanism can be used to slow down TCP traffic rate if congestion develops in the intermediate nodes. The ELN mechanism can be used to retransmit lost packet instead of reducing the window size. If packet loss occurs due to the route failure, then TCP-F or ATCP scheme can be used to deal with this problem. In the presence of variable SNR, adaptive modulation technique can be used to improve the network throughput. In addition, link prediction methods can be used to estimate the SNR based on the previous measurements in order to change the route before the SNR reaches a pre-defined threshold, hence packet drop can be prevented.

3.3 Architectures for implementing cross layer design

In the TCP/IP layered model, layers communicate with each other by setting a specific field in the header [16]. For example, in the standard TCP/IP model as shown in figure 3.10, the data link layer knows which protocol is running on top of it by looking in the “type” field. The network layer knows about the protocols at the transport layer through the “protocol” field in the IP header. Similarly, the transport layer knows about the ports at the session layer through “ports” field in the TCP or User Datagram Protocol (UDP) header.

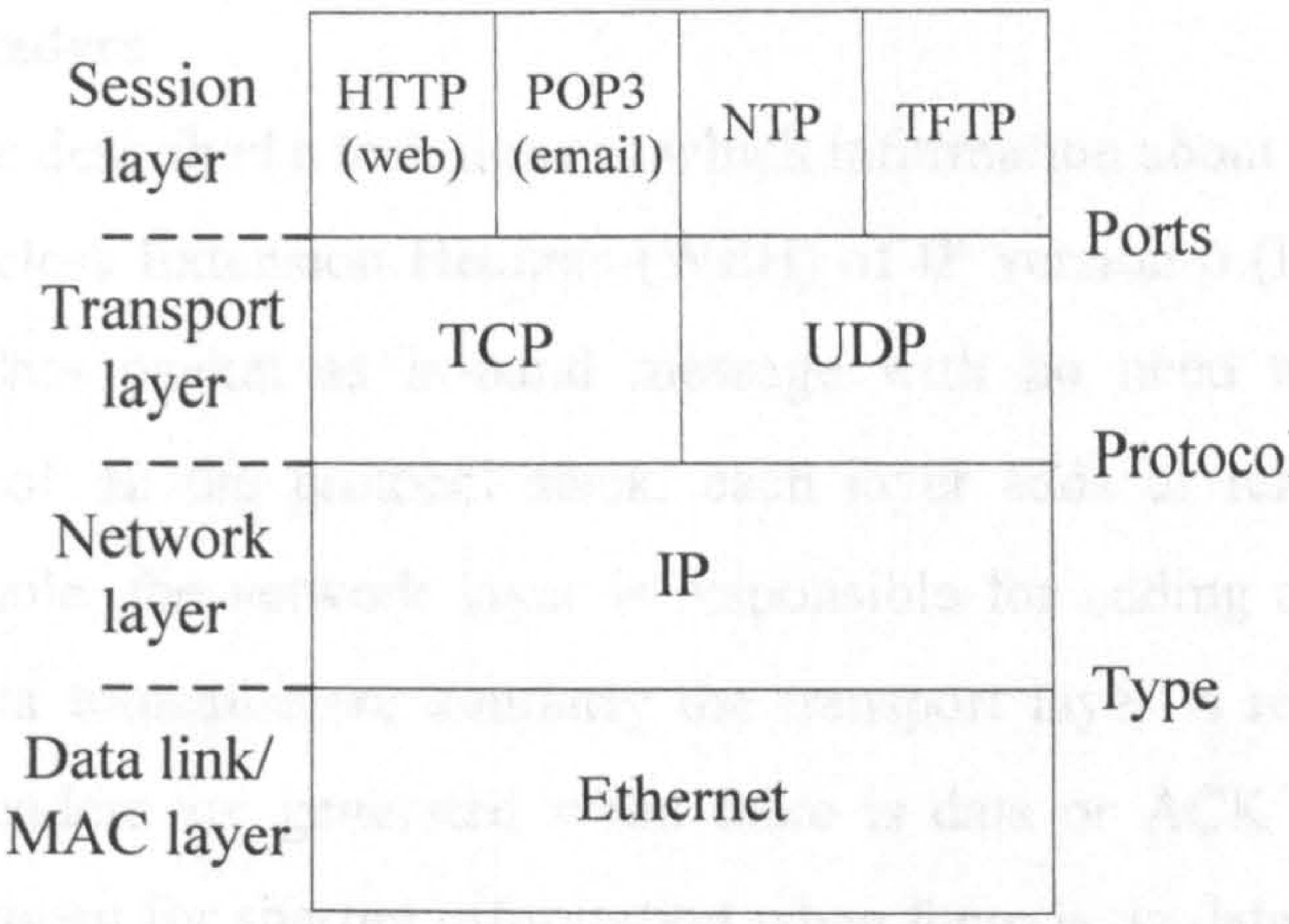


Figure 3.10 Interactions between the layers in the layered model

The layered approach is an important feature that allows communication protocols and standard interfaces to be used, however as discussed in the section 3.1.1 that it is not well suited for wireless networks. In section 3.2, potential cross layer solutions have been proposed. This section discusses the different possible cross layer architectures. The cross layer architecture defines how variables of one layer can interact with other layers in the network model. Several methods for implementing cross layer communication have been proposed in the literature (see figure 3.11) which are discussed below.

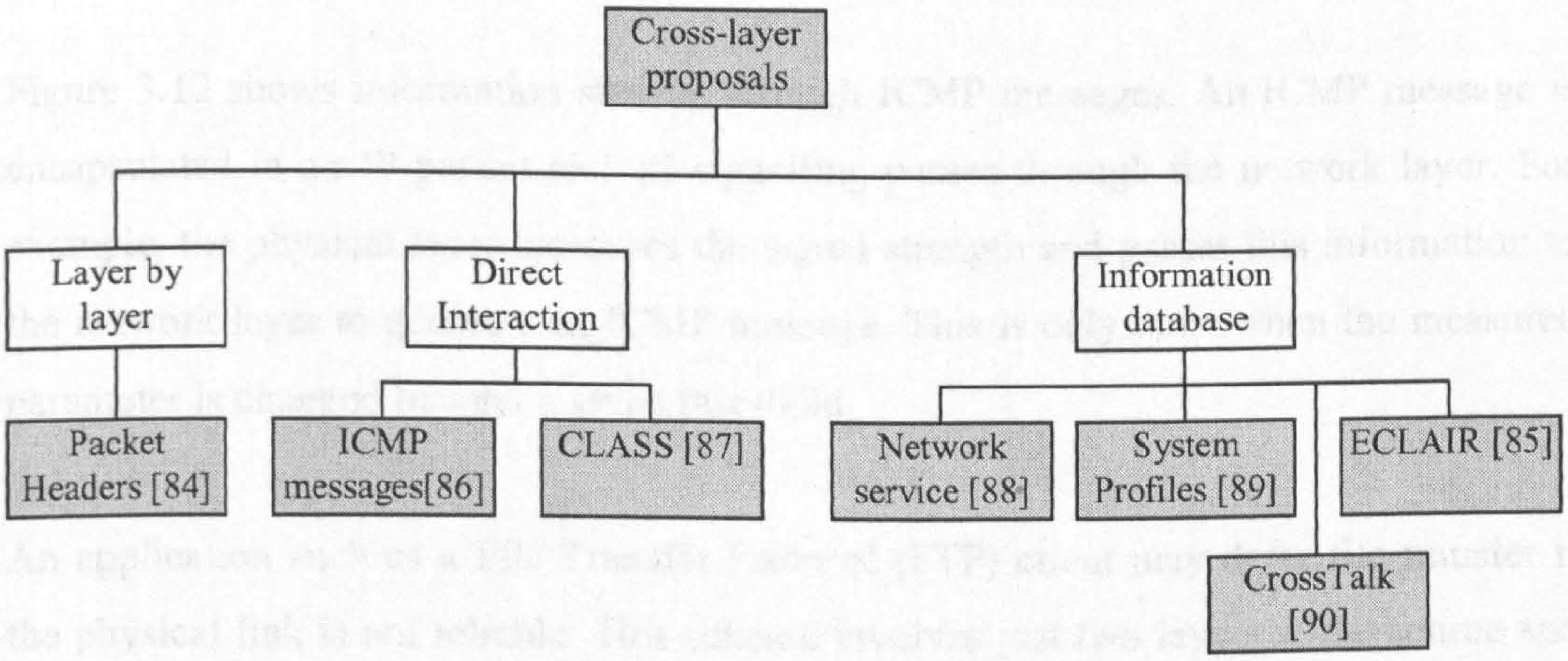


Figure 3.11 Cross-layer design proposals

3.3.1 Layer by layer proposal

3.3.1.1 Packet headers

Wu et al [84] have described a technique in which information about the other layers are added in the Wireless Extension Headers (WEH) of IP version 6 (IPv6) packets. This method uses a data packet as in-band message with no need to use a dedicated messaging protocol. In the protocol stack, each layer adds or removes the relevant header. For example, the network layer is responsible for adding or removing the IP header during data transmission; similarly the transport layer is responsible for TCP header. Packet headers are generated when there is data or ACK to send. However, there is no mechanism for sharing information when there is no data transmission. This method is used for interlayer signalling within a host; however this is more suitable for external IP-level information interchange.

3.3.2 Direct interaction proposal

3.3.2.1 ICMP messages

ICMP as described in RFC 791 [91] is sent when an IP datagram cannot reach its destination, or there is not enough buffer capacity at the router. ICMP works on the top of the IP layer and provides a feedback mechanism for an unreliable IP service. Sudame et al [86] use ICMP messages which propagate information such as signal strength, handoff, available energy, and bandwidth in the network.

Figure 3.12 shows information sharing through ICMP messages. An ICMP message is encapsulated in an IP packet and all signalling passes through the network layer. For example, the physical layer measures the signal strength and passes this information to the network layer to generate an ICMP message. This is only done when the measured parameter is changed beyond a given threshold.

An application such as a File Transfer Protocol (FTP) client may defer file transfer if the physical link is not reliable. This scheme involves just two layers at the source and destination nodes when any change of parameter occurs as opposed to the packet header scheme where information has to propagate layer by layer. In addition, information sharing between the layers is independent of data transmission; hence layers have knowledge about the current network status before the data transmission, and can adapt without adding delay.

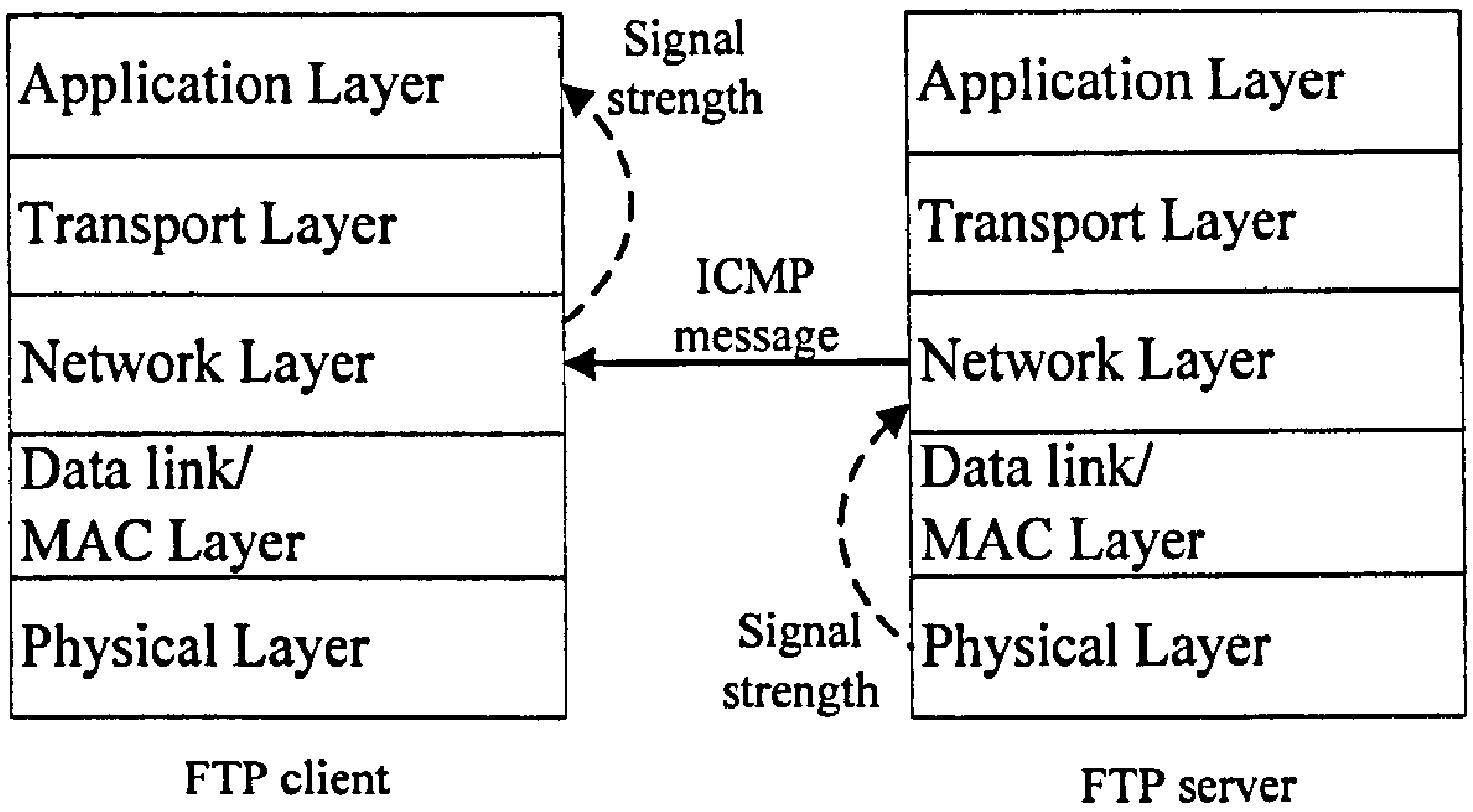


Figure 3.12 ICMP based architecture

3.3.2.2 Cross Layer Signalling Shortcuts (CLASS)

Wang et al [87] have proposed a scheme in which layers can directly interact with each other through a dedicated internal message. The communication between the layers is achieved through sending a 4 byte message to the destination layer, see figure 3.13. The message format consists of the following fields:

Destination Address (1-byte): Includes destination layer, destination protocol or application.

Event Type (1-byte): Indicates a parameter.

Event Contents (2-bytes): The value of the parameter.

The proposed internal message is used to signal between the layers within the mobile host, however signalling between the mobile and other hosts can be done through signalling protocols such as ICMP.

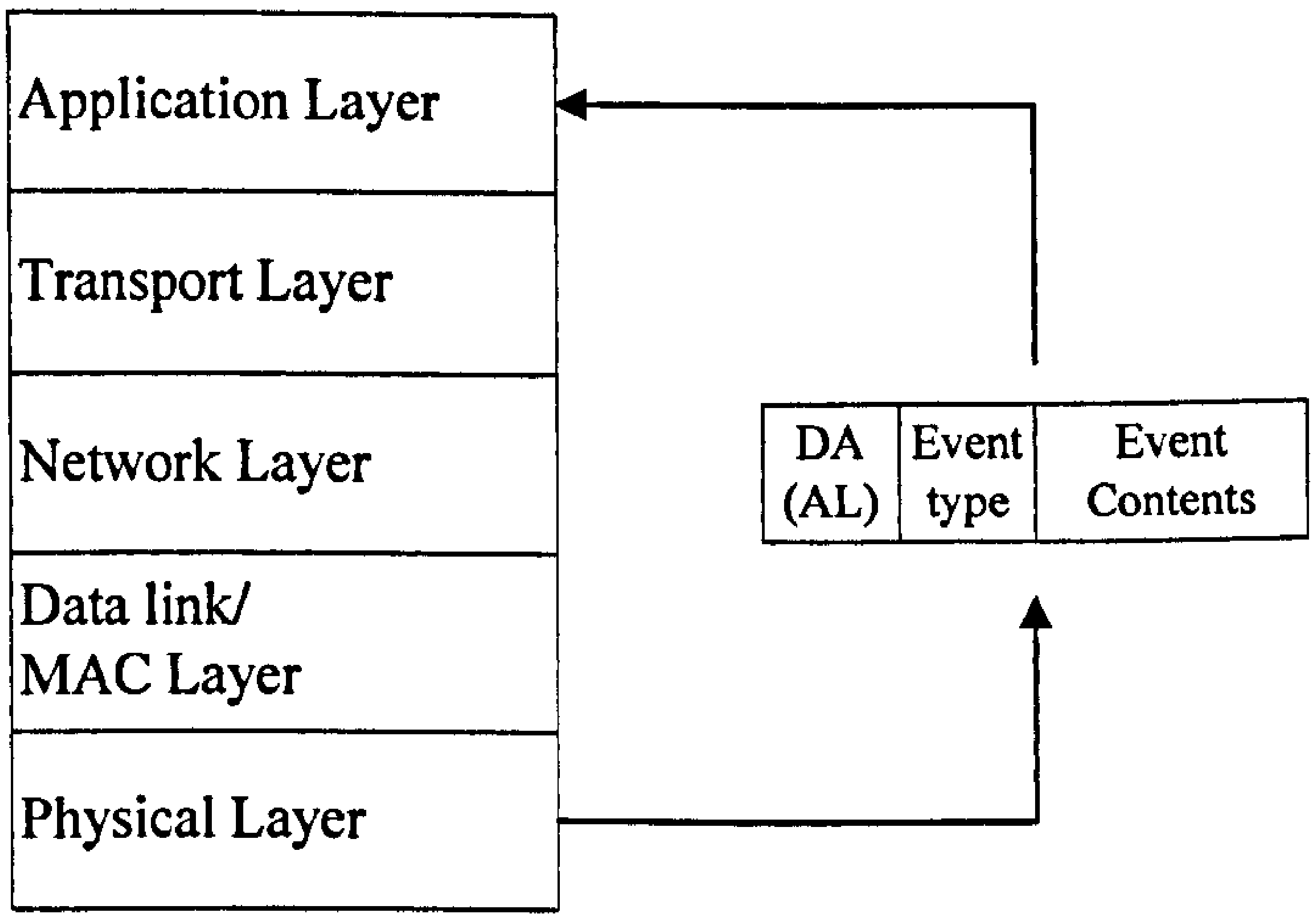


Figure 3.13 Interlayer communication through internal message

The benefit of this scheme as compared to the ICMP scheme is the reduction in the overhead; CLASS uses 4 bytes of internal message whereas common ICMP header is 8 bytes long.

3.3.3 Information database

3.3.3.1 Network service

In this scheme, information sharing between the layers is managed by the Wireless Channel Information (WCI) server [88]. The WCI collects and processes physical/MAC (PHY/MAC) layer parameters from the base or mobile stations over the air interface such as SNR, MAC frame and payload size. The WCI then processed these parameters and calculates available bandwidth, latency, and packet error rate. These parameters are then sent to the application server. The application running at the mobile station can adapt with the help of information provided by the application server.

Figure 3.14 shows how the PHY/MAC layer parameters are transmitted to the WCI server over the air interface. The WCI estimates the average bandwidth (bps) over a specified time period. This is estimated by monitoring a continuous stream of packets that are correctly delivered over a specified period of time. The latency is measured which is the average time required for reliable delivering an IP packet of a certain length. The WCI estimates the packet error rate of a given IP packet length. Communication between the WCI server and the application server is over the air interface. Hypertext Transport Protocol (HTTP) over TCP is proposed for infrequent exchange of parameters. However, in the case of a rapid exchange of short messages where parameters have a short life span, TCP over UDP is suggested because the retransmission mechanism of a reliable transport protocol like TCP may no longer be meaningful after a delay.

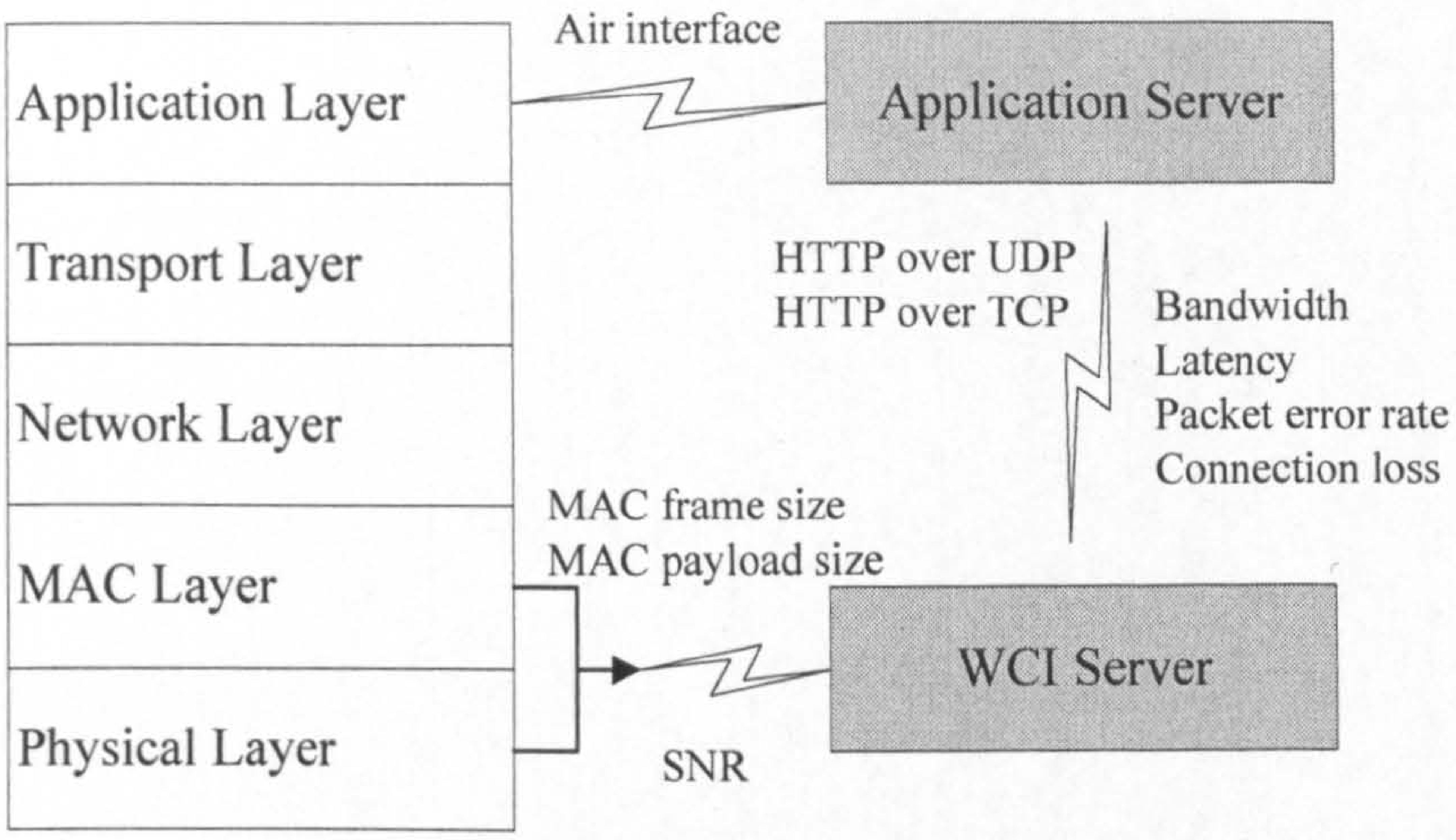


Figure 3.14 Network service scheme

The scheme avoids processing overhead of the PHY/MAC parameters at the application layer because such parameters are processed outside of the layered stack. However, this scheme is applicable where WCI and application servers are fixed in the network and hear other base and mobile stations. In addition, minimum delay is needed when the information is propagating through the air medium; otherwise the adaptation at the application layer may no longer applicable.

3.3.3.2 System profiles

Chen et al [89] describe a scheme in which information sharing between the layers is managed through system profiles. Each layer provides abstracted information about the node and this is stored in the system profiles within the node. Other interested layers can then select the profile to fetch the desired information. For example, in figure 3.15, the application layer sets priority, data quality level, and loss tolerance as an application-level requirement, and then translates such a requirement into a set of network-level QoS parameters such as bandwidth, delay, and reliability. Compared to the network service scheme where information sharing between the layers is performed through external servers; system profiles are saved locally within the node and information between the layers can be shared rapidly.

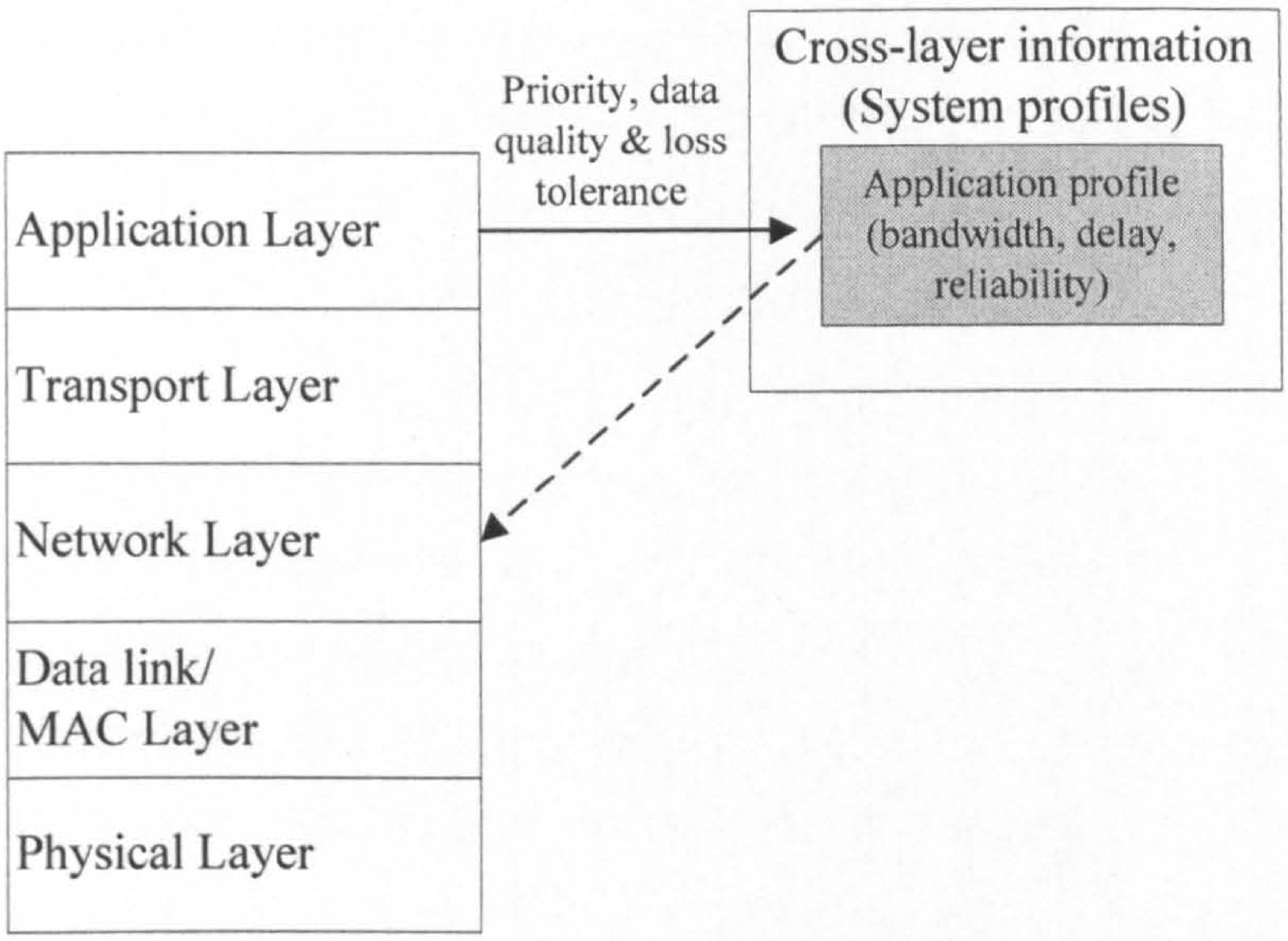


Figure 3.15 System profiles scheme

3.3.3.3 ECLAIR architecture

Raisinghani et al [85] have proposed a guideline for designing and implementing a cross layer feedback architecture at the user end devices. Layers can interact with each other through a protocol optimizer (PO). The design uses a tuning layer (TL) which provides an interface between the layered protocols and PO, see figure 3.16. The PO optimizes protocol functions based on the user defined algorithm. For example, a user application sets the maximum delay that the application can tolerate. The algorithm for the IP routing protocol optimization takes the maximum delay from the Application Tuning Layer (ATL) and makes it available to the network layer, so that the route can be selected based on the required delay, and hence the routing protocol operation can be optimized.

Another example is TCP optimization. A user may want to download a file and requires a certain throughput. The receiver's TCP window size is notified by the advertised window field in the acknowledgements to the sender. The receiver TCP window size is extracted through the TCP Tuning Layer (TCPTL) which is used to optimize the sender TCP protocol, and hence the send rate of a TCP sender can be determined. With the knowledge of available end to end throughput, the sender is able to use the application with the required throughput.

The author further discusses two types of adaptations; synchronous and asynchronous adaptations. In synchronous adaptation, whenever a layer receives some cross layer feedback, it proceeds with its regular execution only after executing the adaptation required. For example, the network layer discovers multiple routes to the destination node, and then finds the best route based on the required delay. However, in the mean time if the network layer receives another update about a broken route, such information will not be included during the search of the best route. In the asynchronous case, the best route searching is updated on the arrival of new routing updates. Hence asynchronous adaptation provides accuracy while optimizing the protocol performance, and therefore this approach is used in the ECLAIR architecture.

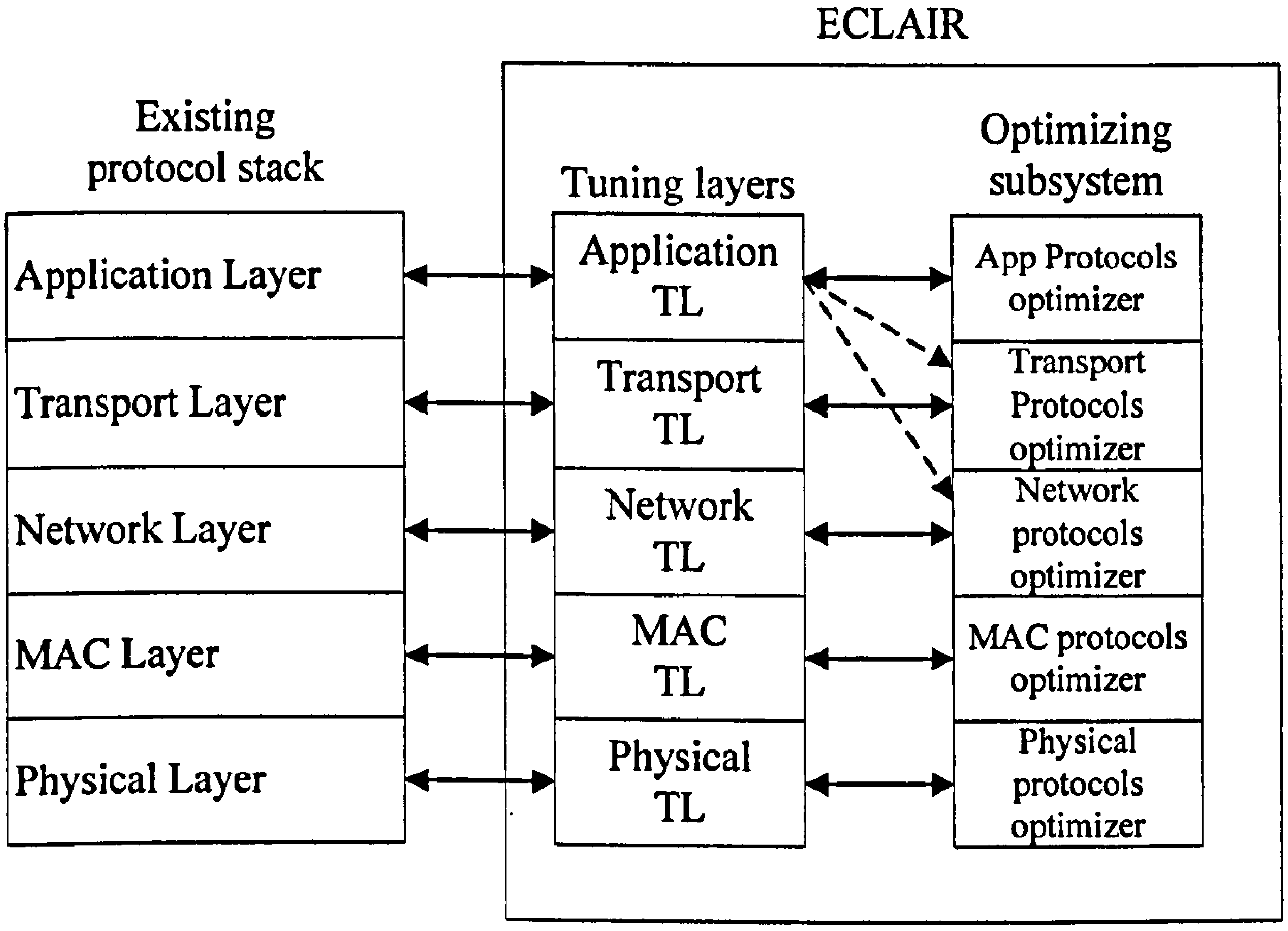


Figure 3.16 ECLAIR: cross layer feedback architecture

3.3.3.4 CrossTalk architecture

Winter et al [90] propose a cross layer architecture in which each node maintains a local view and global view of the network, see figure 3.17. Local view information is related to the node itself, for example, current battery status, load, SNR, transmit power, location information or velocity. Protocols can also access such data to optimise their local operation. Global view information is related to the information of the network which is based on single or multiple metrics, such as the energy levels of the nodes, the communication load or the number of hops. Global view information of the network at the node is collected with the help of other node's local view information. Local view information is shared by piggybacking in the data packets. Only the source node adds local view information. Forwarding nodes only extract the source node's local information and add in its global information database. However, forwarding nodes do not add any additional information. Similarly, the destination node extracts the source node's local information in the same way as forwarding nodes.

Load balancing is one of the examples which can use this architecture. For example, assume the AODV routing protocol is running in an ad hoc network. Each node adds "load" (local) information in the data packet to inform other nodes that the node is busy forwarding data. Load information is calculated based on the number of RREQ/RREP packets. When a node generates a RREP packet, then the node is selected as a forwarding node for the selected route. On the receipt of "load" information, other nodes are then able to select nodes that are currently handling less communication traffic. The benefit of this scheme is that data packets are used as a signalling mechanism to distribute other node's information, hence another signalling mechanism like control packets is not needed, and signalling overhead can be minimised.

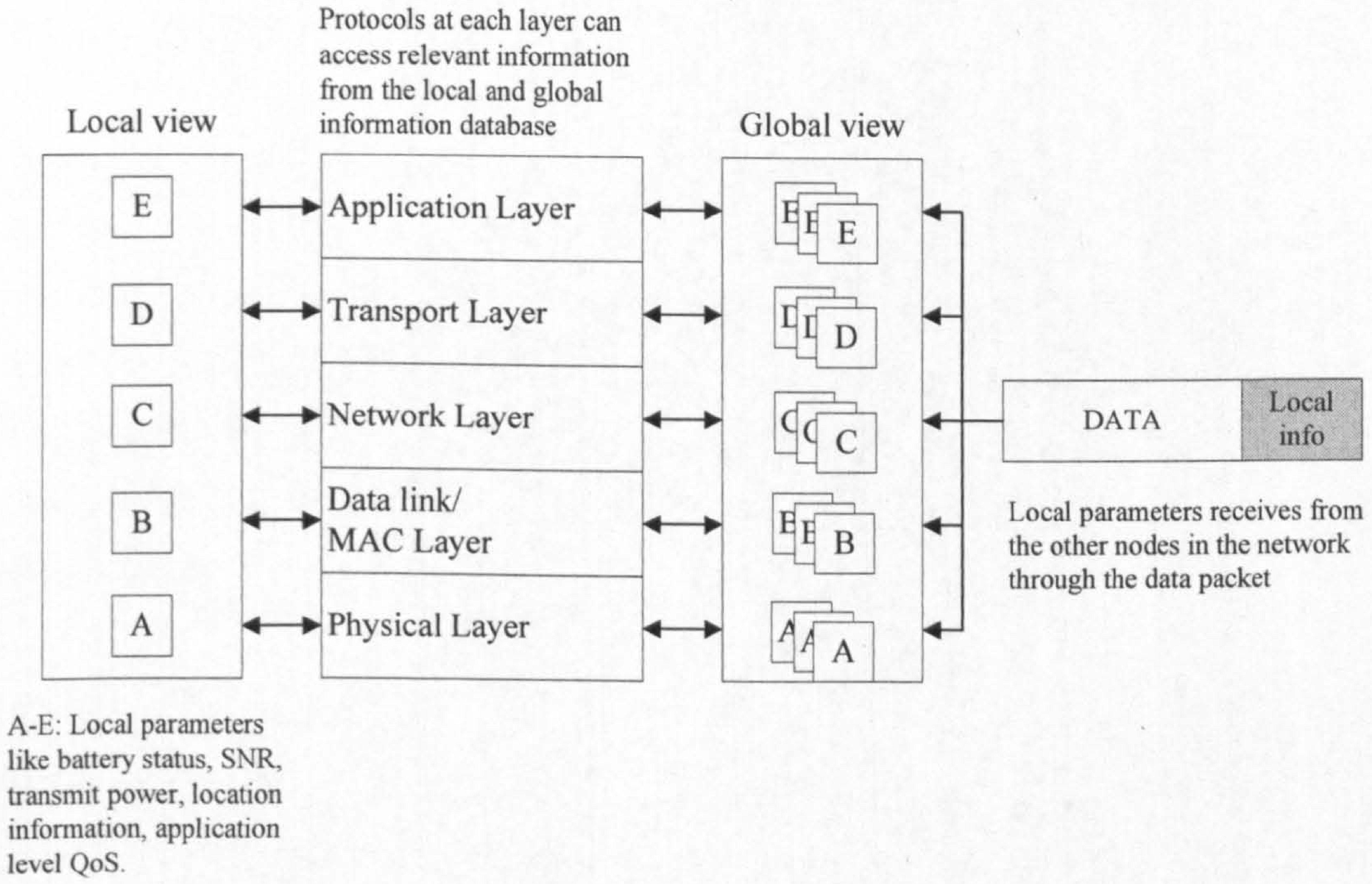


Figure 3.17 CrossTalk architecture

3.4 Cross layer architecture for ad hoc networks

3.4.1 Requirement for a multi hop ad hoc network

Selecting a cross layer architecture depends how frequent information needs to interchange. For example, in a fixed ad hoc network, if a route failure happens due to battery depletion in a relay node, then the physical layer shares this information with the network layer to change the route; and with the transport layer to halt TCP transmissions. Route breakage events due to battery depletion may not be frequent. However, when nodes are mobile then a route failure could occur due to relay nodes leaving the network, or due to high attenuation. In this case, the physical layer frequently sends information to the network and transport layer to adjust and adapt their protocols accordingly.

3.4.2 Dynamics of cross layer operation

When time sensitive information is passed between the layers, the validity of such information is important. For example, in section 3.2.4, the application layer defines the required modulation to the physical layer based on the SNR value at the receiver end. However there are two conditions:

a) Intermittent change: If the SNR value changes slowly, then during the process of selecting a suitable modulation, an external module should monitor the SNR during the selection of modulation, and if the SNR value changes significantly then it should cancel the current modulation selection process by sending a signal to the application and physical layers and restart the process. This condition only exists if the variables which are used to adapt the protocols at different layers change infrequently but significantly.

b) Continuous change: If the SNR value changes rapidly, then the selected modulation based on the SNR may no longer be valid. In order to avoid inaccurate feedback, modulation based on the lowest SNR value can be used. However lower order modulation may not provide the required throughput for the application.

3.4.3 Optimum solution and architecture

An optimum cross layer architecture depends on how rapidly information is required to be shared between the layers. Inter-layer communication through protocol headers [84] adds information in the header of the data packet to communicate with other layers. The information propagates layer by layer which adds delay in getting information from the originating layer to the destination layer. For example, if the physical layer sends information to the TCP layer to resume suspended transmission, by the time it reaches to the TCP layer, the signal quality may deteriorate. This situation could occur in the ad hoc network, particularly if the nodes are mobile.

A direct interaction method such as the CLASS scheme [87], uses a dedicated message to share information between two layers. However, if the information is required by more than two layers such as when the signal is experiencing deep fades for a length of time, then the physical layer notifies the routing layer to change the route for future data transmission, and also notifies the transport layer to halt TCP transmission to avoid data loss. Hence two messages are required, and in the case when a single parameter affects more than three layers, this increases the frequency of internal messages and the load on each layer to generate such messages. In addition, the load increases if parameters are affected by the mobility of the nodes.

Another direct interaction proposal, the ICMP scheme [86] also has a similar problem as the CLASS scheme. Continuously changing parameters will increase the frequency of ICMP messages and generates more signalling load on a layered architecture.

An information database proposal such as network service [88] uses a Wireless Channel Information (WCI) server to share information between the layers. WCI exist physically in the network as a dedicated server, which is not possible in an ad hoc network, which does not include any fixed infrastructure.

The other database proposals such as ECLAIR [85] and CrossTalk [90] can be used in an ad hoc network. However these schemes differ in their means of distributing information between the other nodes in the network. The ECLAIR architecture uses acknowledgement, however the CrossTalk architecture disseminates information through the data packets. The feedback mechanism depends on how quickly the information is required by other nodes. For example, SNR value changes with the speed of the nodes, hence the frequency of broadcasting SNR information should relate to the speed of the nodes, otherwise the source node may not be able to adjust the transmission rate correctly due to outdated SNR information. A separate control channel is needed for this purpose. On the other hand, battery energy information does not change rapidly with time, and this information can be added in the data transmissions or acknowledgements.

3.5 Summary

A layered model was originally developed for a fixed network. In wireless networks, variations at one layer have an impact on other layers (layers are not independent).

Network performance can be improved by using cross layer interactions. For example, sharing information between the network layer and transport layers can prevent TCP retransmissions when a route fails. The application layer shares SNR information from the physical layer and decides which modulation to use based on the required application throughput and informs the physical layer to use selected modulation. The physical and network layers share SNR information so that at low SNR value, the network layer changes to a new route before the SNR falls below a pre-defined threshold. If SNR value is below the pre-defined threshold for some time, then the SNR value is shared with network, transport and application layers to halt data transmission for the destination node.

There are several techniques in which information can be shared between different layers. Some techniques use a layer by layer approach to share information between the layers (using packet headers); some use direct interaction (through dedicated messages); or some use an information database (through external servers or system internal sharable memory).

The method selected to transmit data between layers depends upon how frequently information needs to be shared and whether information is time sensitive or insensitive. Information sharing between the layers using an information database through system internal sharable memory is considered the best approach in a wireless network because information in the database is accessible to all the layers.

4. Maximising the Network Lifetime

4.1 Introduction

An ad hoc network is a collection of nodes that form a full or partial mesh. Routes between the two nodes can be formed either directly or with the help of other relay nodes. The nodes in an ad hoc network are generally battery powered and so the operation time of the nodes is limited by their battery charge. There are two important criteria:

i) Minimising the power consumption of individual nodes

ii) Maximising the time over which all nodes are available for communication.

It may not be possible to satisfy both of these criteria simultaneously [15]. For example, if the network frequently selects the same nodes to act as relay nodes on the basis of power minimisation, then the battery charge of these nodes will be rapidly depleted, and therefore reduce the time when all nodes are available. This will not only eliminate the depleted nodes from the network but the lack of relay nodes may also prevent other nodes from communicating. Therefore, one of the most important and challenging issues in these networks is how to balance the use of nodes to maximise the time when all nodes in the network can communicate. The multi-route capability of an ad hoc network gives a degree of flexibility in the choice of nodes that can be used in this balancing process.

Minimum power routing and maximum lifetime routing schemes are discussed in the following sections.

4.2 Minimum power routing

The aim of a minimum power routing scheme is to minimise the total power consumed in forwarding data from a source to a destination. The power consumed by the source depends on the distance and clutter (for example attenuation due to objects in the transmission path) between the source and destination nodes. If the source node dynamically adjusts its transmission power so that the transmitted signal is just strong enough to reach the destination node, it minimises the power needed to transport data to the intended destination. In some circumstances, the power can be reduced if data is transmitted through one or more relay nodes [28, 92, 93].

Adding relays between the source-destination pair may save power because the received power is inversely proportional to the $1/d^n$, where 'd' is the distance between the nodes, and 'n' is the path loss exponent [94]. If a path is divided into a number of short hops then the power budget on each hop is improved corresponding to the reduction in path loss, and this can be exchanged for a reduction in RF transmitter power. Therefore using a route with a large number of small hops may be able to provide a lower power solution than a route with small number of relays. However for each node used in the route, power is used not only for the RF transmission but also for receiving and processing; as the number of hops increases, the increase in power consumption due to processing outweighs the reduction in transmit power. In [95], two multi hop models (nodes with equal spacing and without equal spacing) are discussed and the power is calculated as a function of the number of hops relative to a single hop route. When the processing power is 5% of the RF power, then the reduction in power is up to 6dB for three hops depending on the position of the nodes. However, when processing is reduced to 1% of the RF power, then the reduction in power is up to 12dB for three hops.

4.3 Maximum network lifetime schemes – literature review

An algorithm selecting the minimum power route may lead to overuse of certain nodes which will lead to early depletion. In order to avoid early depletion, traffic should be routed through nodes that have sufficient remaining battery. However, the route may not be the minimum power route.

4.3.1 Power, residual capacity and residual battery energy routing

Several power aware routing schemes have been proposed to conserve battery. Michail et al [96] propose a scheme where the objective is to maximise the amount of information transmitted per energy unit while balancing energy usage among all the network nodes and uses three metrics. The first metric (M1) is a direct measure of the power needed for transmission between the nodes, and selects the minimum end to end power route. Hence, the cost (D_i) of this metric is:

$$D_i^{(1)} = \min \sum_{i=0}^{N-1} P_i \tag{4.1}$$

where

i=0 Source node

N Total number of nodes in the route; N-1 is the number of relay nodes in the route

P_i Transmission power of node i

M1 will always route traffic through the minimum power path. However, when the minimum power route is used frequently in the network, then some nodes get heavily utilized and some remain lightly used. In a static topology, the minimum power path will be the same until any node's battery charge reaches zero. In order to avoid the heavily used node, a second metric (M2) is proposed which attempts to discourage nodes which have low small available capacity. The available capacity of node i is defined as the number of available communication transceivers (R_i) that can support up to R_i sessions simultaneously. The number of reserved transceivers (B_i) varies with time in the network. The available capacity of node i (number of free transceivers) is calculated as: $R_i = C_i - B_i$, where C_i is the total number of transceivers in node i.

The cost of M2 metric is calculated as:

$$D_i^{(2)} = \min \sum_{i=0}^{N-1} \left[\frac{P_i}{R_i} \right] \tag{4.2}$$

where

R_i available communication transceivers in node i

The drawback of the first and second metrics is that the node's residual battery energy is not considered during route selection, and some nodes may be more heavily used than others. The third metric (M3) is proposed which attempts to introduce some fairness considerations in node usage, so that the energy is evenly consumed by all the nodes in the network. The M3 metric takes the residual battery energy of each node into account and discourages routes which have a low residual battery energy; it also includes a minimum power routing metric. Weights are given to these metrics, so that both of the metrics can be used. When the network is first used, all nodes have full battery capacity; hence the route is selected on the basis of the first metric (M1). As the residual energy of the used node begins to drop, the cost of links with a low battery node becomes high; and the residual energy metric weight is more dominant than minimum power metric weight. The cost of the third metric is calculated according to the following equation:

$$D_i^{(3)} = \min \sum_{i=0}^{N-1} \left[\left(W_p \times \frac{P_i}{P_{\max}} \right) + \left(W_e \times \frac{E_i^o}{E_i^R} \right) \right] \tag{4.3}$$

where

W_p & W_e	weighting factors
E_i^o	initial amount of energy available to node i
E_i^R	residual amount of energy available to node i
$P_{\max}$	maximum transmission power level of a node ($P \leq P_i \leq P_{max}$)

W_p and W_e are the weighting factors in which two cases are discussed; a) residual energy only (Wp=0 and We=1), and b) transmission power and residual energy (Wp=1, We=1). In the beginning of the network, the second term for $D_{ij}^{(3)}$ is one for all the nodes. With time, the second term will increase and the amount of residual energy of nodes is low which increases the cost of the route.

It is assumed that each node maintains up to date information about the identities of next hop neighbours, required transmission power, residual capacity and energy. All such information is broadcasted periodically to other nodes that are located within transmission range. Each node is equipped with multiple transceivers however a multiple access technique such as Space Division Multiple Access (SDMA), Frequency Division Multiple Access, (FDMA), Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) is needed for simultaneous communication. In the wireless network, connectivity between the nodes is based on the distance between the nodes, transmission power and minimum required receive power. In addition, there are no obstacles and therefore all nodes are reachable within communication range of the transmitting node. Metrics 1 and 2 use a lower aggregate power route and therefore achieve a lower average energy per route. Metric 3 uses power and residual energy. This comes at the cost of an increase in average energy per route. The paper does not analyse the effect of these metrics on the network lifetime.

4.3.2 Power-aware source routing

Maleki et al [97] present a Power-Aware Source Routing (PSR) protocol which extends the network lifetime by balancing node usage. The PSR is based on the Dynamic Source Routing (DSR) protocol [37] which uses RREQ/RREP packets to find the route from the source to the destination node. It is assumed that all the transmission power levels in PSR are constant. In this scheme, the source node calculates the link cost $C_i(t)$ according to the following equation:

$$C_i(t) = \min \sum_{i \in Z} \rho_i \left(\frac{F_i}{R_i(t)} \right)^{\alpha} \tag{4.4}$$

where

ρ_i	transmit power of node *i*
F_i	full-charge battery capacity of node *i*
R_i	remaining battery capacity of node *i* at time *t*
α	a positive weighting factor
Z	(i, a, b, c ... j), i = source node, j = destination node, and a, b, c are intermediate nodes

The link cost is added in the header of RREQ packet and broadcast in the network. Intermediate node receives RREQ packet, calculate link cost and waits for additional RREQs. If a new RREQ arrives with the same destination and sequence number, it compares the cost of the newly arrived RREQ and compares it with the previously arrived RREQ link cost. The RREQ which has a minimum link cost is added in the header of the RREQ packet and broadcast in the network. The destination node receives several copies of the RREQ packets which show the different possible routes from the source node. The destination node selects the route with the minimum cost and sends a RREP. When the energy of the node is low, the cost for routing a message through that node is high, and hence the network load is balanced by avoiding routes through low-energy nodes. The network lifetime is defined as the "time taken for K% of the nodes in a network to die". Lifetime is compared with DSR and it shows that PSR gives an improvement in the network lifetime of each node. This is due to the fact that DSR selects route on the basis of minimum hops without considering battery and the possibility that some nodes can be used more than others. PSR selects a route on the basis of battery utilization and the node's battery is evenly utilized.

The increase in the network lifetime does not consider the processing power of the nodes. The routes are selected on the basis of node's battery utilization. The PSR approach looks into the lifetime of the node without conserving transmitter power since the transmitter power is constant. The authors did not discuss the value of K and how does it varies with the number of nodes, network size, and the effect on the reachability to the destination nodes.

4.3.3 Maximum survivability routing

Marbukh et al [98] present a Maximum Survivability Routing (MSR) which is aimed at prolonging the network connectivity by choosing routing paths according to the remaining battery life of the nodes along the route. In this approach, each node calculates life expectancy with the help of the power draining rate according to the following equation:

$$r_i(t) = \frac{P_i(0) - P_i(t)}{t} \tag{4.5}$$

where

$r_i(t)$ battery power draining rate of node i at time t

$P_i(0)$ full-charge battery capacity of node i

$P_i(t)$ remaining battery capacity of node i at time t

The power draining rate is used to calculate the life expectancy of node i as:

$$T_i = P_i(t) / r_i(t) \tag{4.6}$$

where

$P_i(t)$ transmit power of node i at time t

Life expectancy is used to compute link cost (between node i and node j) according to the following equations:

$$C_{ij} = \min \sum_{i \in R} \left(\frac{1}{T_i} \right) \tag{4.7}$$

where

R = (i, a, b, c … j), i = source node, j = destination node, and a, b, c are intermediate nodes.

The link cost is distributed to all nodes in the network via data and control packets periodically. A dynamic routing table is maintained by each node. For each destination, a node stores the preferred route which is based on the cost function.

The author assumes that all nodes are reachable, and therefore nodes can receive link cost information from all the nodes. Such information can be blocked if there are obstacles in the network. As a result, some destination nodes may be unreachable or reachable via a non-preferred route such as via low battery relay nodes. It is assumed that each node has knowledge of transmission power to communicate with other nodes, however the mechanism for achieving this is not discussed. Routes are selected which consist of strong battery nodes without conserving the end to end power consumption of the route.

4.3.4 Conditional max-min battery capacity routing

Toh [15] has proposed a conditional max-min battery capacity routing (CMMBCR) algorithm. The algorithm uses a minimum power routing until the remaining battery charge reduces to a pre-selected threshold. When this threshold is reached, the node is withdrawn from use as a relay node which protects individual nodes from overuse. The performance of the CMMBCR routing depends on the chosen threshold. Toh's investigation reveals that when the threshold value increases, the time to the first node failure increases. Due to early isolation of relay nodes from the network, other nodes in the network may use a higher transmission power or longer hop routes in order to communicate. As a result, the battery consumption of other nodes increases. On the contrary, if the loss of a few nodes is acceptable in the network, then a lower threshold value can be used, and hence this saves the battery consumption of the remaining nodes.

4.4 Analysis and optimisation of the CMMBCR scheme

The network simulations reported by Toh [15] for the CMMBCR scheme calculate the times at which the batteries in each node in the network become depleted. It also shows how these times vary with the value of the threshold selected for withdrawing nodes from use as a relay node. It does not, however, show how the battery charge reduces over time or how the use of a threshold changes the probability that a node is going to be used as relay. Also it does not identify the optimum threshold level for the maximum network lifetime.

This chapter carries out a more detailed analysis to explore the evolution of the battery charge with time and hence to identify the optimum threshold.

For many applications, it is important to ensure that all users are able to communicate in a network. Therefore, network lifetime is defined as the time until the first node reaches zero battery charge.

The rate at which current is drawn from the battery in an ad hoc network node depends on the following:

- The RF power needed to transmit to the next node which depends on the link length and attenuation due to clutter.
- The efficiency of the RF power amplifier.
- The power required for signal processing and receiving.
- The probability of a node being used as a source and also as a relay.
- The percentage of time a node is transmitting.

WLAN handsets use a Class-AB RF power amplifier [99]. The maximum RF power for wireless handsets is specified by the standardization bodies; for WLAN handsets, a maximum RF output power of 100mW is specified [100, 101]. The current used is dependent on the efficiency of the power amplifier, which is defined as the ratio of RF output power to the electrical input power. The maximum efficiency of the class AB amplifier is about 40% [102].

4.4.1 Investigation of battery usage with minimum power routing (MPR)

In order to analyse the battery usage with and without a threshold, a simple analytical model of 25 nodes has been used. Figure 4.1 shows an example of an ad hoc network including areas of attenuation, or clutter that represents obstructions such as buildings. Routes between nodes are selected using relays as required, based on achieving the minimum power route from source to destination. It can be seen from this example that the minimum power route between some nodes may be a direct connection whereas for other routes a low power route may be found by using a relay to circumvent clutter. As a result, depending on their location, some nodes have a higher probability of being used than others.

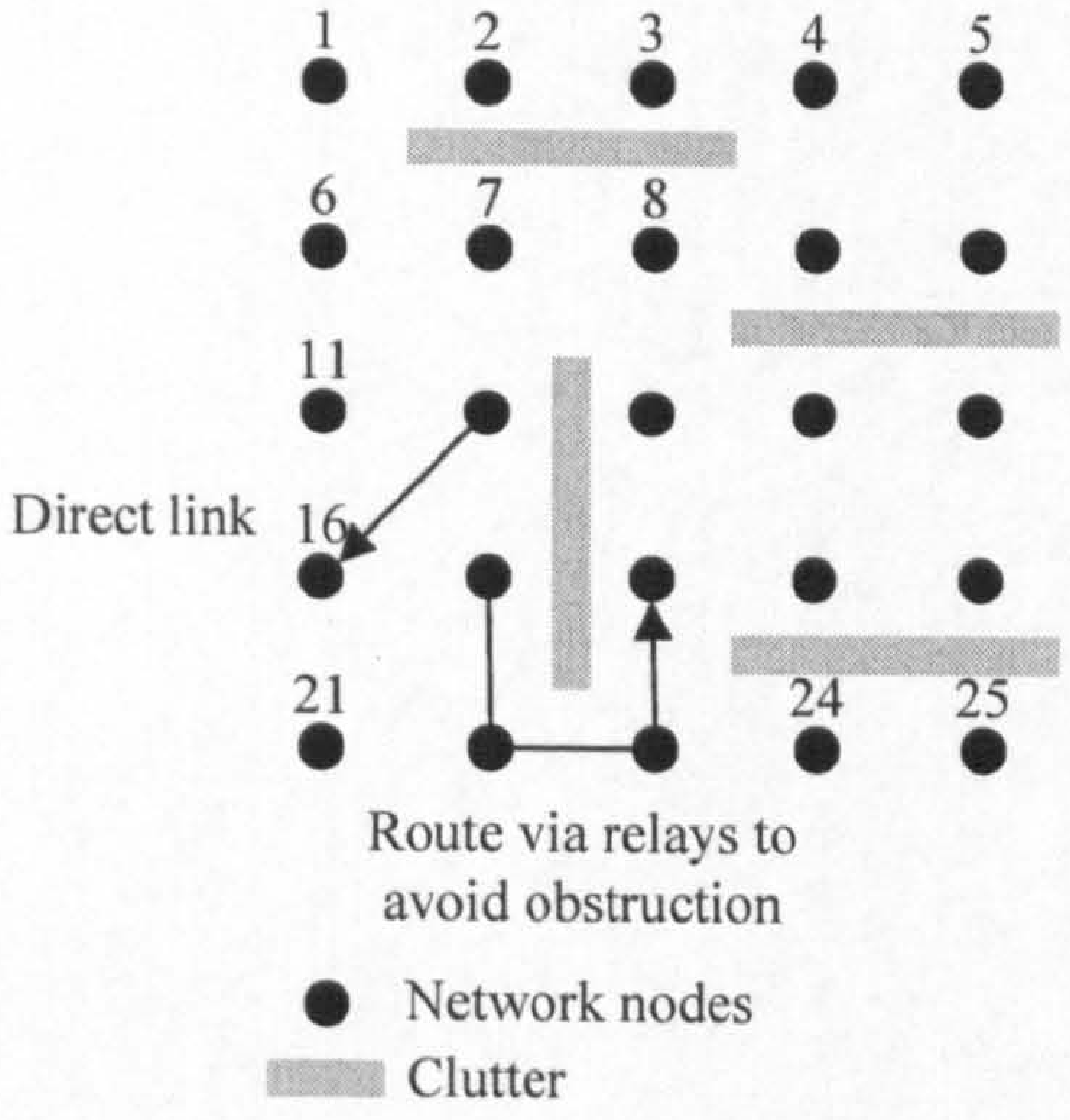


Figure 4.1 Example network

The network lifetime is evaluated based on the following assumptions and parameters:

- WLAN enabled handheld terminals are assumed which have an integrated single chip designed for low power applications. The average current for signal processing when a node is transmitting data is 219mA, and the average current for signal processing when a node is receiving data is 215mA [103].
- There is no limit placed on the maximum transmitter power, hence each node can communicate directly with any other node.
- In the 25-node network model, the maximum transmission power for a point to point link is 5.7W (37dBm). This is based on the specific simulation model defined in figure 6.2. The distance between the nodes is 59m, the clutter factor is 48dB and the fade margin is 11dB. The path loss is 123dB (according to the

equation 5.6 in section 5.4.12). The receiver sensitivity is -86dBm [104]. The transmission power can be calculated with the help of the equation 6.1 in section 6.2.

- Power can be reduced by using relays. The maximum transmission power used by a node in a multi hop link is 347mW (25dBm). This is based on the specific simulation model defined in figure 6.2. The distance between the nodes is 35m, the clutter factor is 45dB and the fade margin is 11dB. The path loss is 111dB (according to the equation 5.6 in section 5.4.12). The receiver sensitivity is -86dBm. The transmission power can be calculated with the help of the equation 6.1 in section 6.2.
- The minimum transmission power used by a node in a multi hop link is 27μW (-16dBm). This is based on the specific simulation model defined in figure 6.2. The distance between the nodes is 7m, the clutter factor is 31dB and the fade margin is 11dB. The path loss is 59dB (according to the equation 5.6 in section 5.4.12). The receiver sensitivity is -86dBm. The transmission power can be calculated with the help of the equation 6.1 in section 6.2.
- The average transmission power is 174mW. The average current used by a node is (174x2.5)/3.3 = 132mA (assuming operating voltage of 3.3volts and RF power amplifier efficiency of 40%).
- There are 25 nodes in the network and 24 possible destination nodes. The total number of routes is 600 (25x24). The probability of a node being used as a source node is 0.04 (1/25). The minimum probability of a node being used as a relay node is 0.003 (2/600). The maximum probability of a node being used as a relay node is 0.04 (24/600). The maximum and minimum usage depends upon the random clutter factor between the nodes in the network.
- The initial battery charge is 700mA-hrs for each node.
- Each node transmits data size of 1MB at 6Mbps. This corresponds to the data transmission time of 1.4 seconds. There is always one node transmitting at one time and nodes are randomly selected.

The following equation predicts the remaining battery charge at time t for node i:

$$BE_i(t) = BE_{0i} - \sum_{D=1}^{t/\Delta t} \left[\begin{array}{l} (I_{SPi(Tx)} \times P_{Si} \times \Delta t_D) + (I_{1i} \times P_{Si} \times \Delta t_D) + \\ (I_{SPi(Tx)} \times P_{Ri} \times \Delta t_D) + (I_{SPi(Rx)} \times P_{Ri} \times \Delta t_D) + (I_{1i} \times P_{Ri} \times \Delta t_D) + \\ (I_{SPi(Rx)} \times P_{Di} \times \Delta t_D) \end{array} \right] \tag{4.8}$$

where

$BE_i(t)$ = Total battery charge at time t for node i

BE_{0i} = Initial battery charge for node i

$I_{SPi(Tx)}$ = Signal processing current of node i due to transmission

$I_{SPi(Rx)}$ = Signal processing current of node i due to reception

P_{Ri} = Probability of node i being used as a relay node

P_{Si} = Probability of node i being used as a source node

P_{Di} = Probability of node i being used as a destination node

I_{1i} = Average current of node i due to transmission

Δt = Time period for data transmission

D = Number of data transmissions ($\Delta t_1, \Delta t_2, \Delta t_3, \ldots \Delta t_{t/\Delta t}$)

Every time a node is used as a source node, the battery charge is reduced due to signal processing current being drawn from the battery until eventually the charge is reduced to zero. In addition, the battery charge is reduced due to transmission current being drawn from the battery until eventually the charge is reduced to zero. When a node is used as a destination node, then the battery charge is reduced due to receive signal processing current being drawn from the battery until eventually the charge is reduced to zero. When a node is used as a relay node, then node functions as a receiver and a transmitter; therefore the battery charge is reduced due to signal processing (transmission and reception) current being drawn from the battery until eventually the charge is reduced to zero.

The power consumption at each node has been evaluated using equation 4.8, assuming that the traffic is uniformly distributed between all the nodes. Figure 4.2 shows the remaining battery charge as a function of time. The curves for the most used node (node 1), the least used node (node 25) and the median used node (node 12) are shown. It can be seen that the variation in the probability of usage leads to a spread in the times taken for the battery charge of each node to reach zero. The lifetime is limited by the most heavily used node with the first node having zero battery charge about 16 hours, while the batteries of other nodes have battery charge remaining.

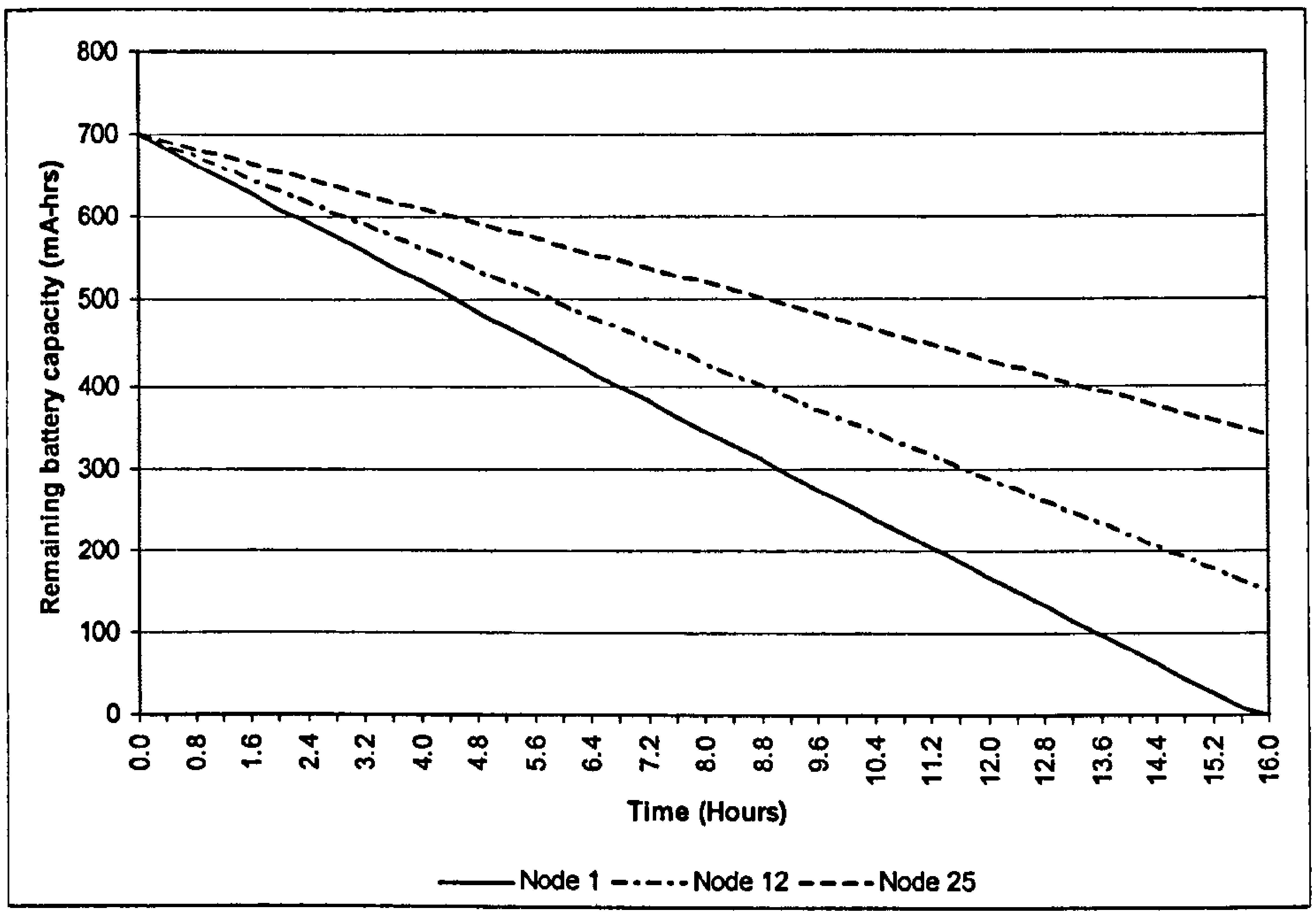


Figure 4.2 Battery charge vs time (Node 1, 12 & 25 – most, median and least used nodes respectively)

4.4.2 Investigation of battery usage using a threshold

When a battery charge threshold is introduced, there are three regions of operation:

a. The first region is where all nodes are available to act as a relay. In this region the rate at which each node's battery is used varies depending on the position of the node in the network. The minimum power route is selected.

b. The second region is the time between the first node reaching threshold and the last node reaching threshold. When a node reaches threshold, it is withdrawn from use as a relay, and therefore minimum power route cannot be selected however, the withdrawn node can be used as a source or destination node. The battery usage of the most used node decreases, but as more nodes reach threshold, the current drain of each node will start to increase as there are fewer nodes available to provide a relay and therefore the RF power required to reach a destination will increase. In addition, the probability of a node selected as a relay node increases.

c. The third region is where all nodes have reached threshold; at this point, relay operation is restored to all nodes and the minimum power route is selected until the first node reaching zero battery charge. The rate of battery depletion in region C is the same as in region A.

The following equation predicts the remaining battery charge at time t for Node i:

When t=0→t=th$_i$ (region A)

$$BE_i(t) = BE_{0i} - \sum_{D=1}^{t/\Delta t} \begin{bmatrix} (I_{SPi(Tx)} \times P_{Si} \times \Delta t_D) + (I_{1i} \times P_{Si} \times \Delta t_D) + \\ (I_{SPi(Tx)} \times P_{Ri} \times \Delta t_D) + (I_{SPi(Rx)} \times P_{Ri} \times \Delta t_D) + (I_{1i} \times P_{Ri} \times \Delta t_D) + \\ (I_{SPi(Rx)} \times P_{Di} \times \Delta t_D) \end{bmatrix} \tag{4.9}$$

When t>th$_A$→t=th$_N$ (region B)

$$BE_i(t) = BE_{0i} - regionA + \sum_{D=th_A/\Delta t}^{(t-th_A)/\Delta t} \begin{bmatrix} (I_{SPi(Tx)} \times P_{Si} \times \Delta t_D) + (I_{2i} \times P_{Si} \times \Delta t_D) + \\ (I_{SPi(Tx)} \times P_{R(th)i} \times \Delta t_D) + (I_{SPi(Rx)} \times P_{R(th)i} \times \Delta t_D) + (I_{2i} \times P_{R(th)i} \times \Delta t_D) + \\ (I_{SPi(Rx)} \times P_{Di} \times \Delta t_D) \end{bmatrix} \tag{4.10}$$

When t>th$_N$→t=EOL$_i$ (region C)

$$BE_i(t) = BE_{0i} - regionA + regionB + \sum_{D=th_N/\Delta t}^{(t-th_N)/\Delta t} \begin{bmatrix} (I_{SPi(Tx)} \times P_{Si} \times \Delta t_D) + (I_{1i} \times P_{Si} \times \Delta t_D) + \\ (I_{SPi(Tx)} \times P_{Ri} \times \Delta t_D) + (I_{SPi(Rx)} \times P_{Ri} \times \Delta t_D) + (I_{1i} \times P_{Ri} \times \Delta t_D) + \\ (I_{SPi(Rx)} \times P_{Di} \times \Delta t_D) \end{bmatrix} \tag{4.11}$$

$BE_i(t)$ = Total battery charge at time t for node i

BE_{0i} = Initial battery charge for node i

$I_{SPi(Tx)}$ = Signal processing current of node i due to transmission

$I_{SPi(Rx)}$ = Signal processing current of node i due to reception

P_{Ri} = Probability of node i being used as a relay node

P_{Si} = Probability of node i being used as a source node

P_{Di} = Probability of node i being used as a destination node

$P_{R(th)i}$ = Probability of node i being used as a relay node (time between the first and the last node reaches threshold)

Δt = Time period for data transmission

D = Number of data transmissions ($\Delta t_1, \Delta t_2, \Delta t_3, \ldots \Delta t_{t/\Delta t}$)

I_{1i} = Average current due to transmission for node i in region A when all nodes are available in the network

I_{2i} = Average current due to transmission for node i after first node has reached threshold and the last node reaches threshold

I_{3i} = Average current due to transmission for node i in region C when all nodes are available in the network

- **Using a low threshold**

The network lifetime is evaluated using the same parameters as described in section 4.4.1. Figure 4.3 shows that nodes are available as relays for a longer time, and the most used node is isolated from relay when only 10% of battery charge left. After that, its usage is reduced to functioning as a source node or destination node. Region B is not clearly visible and region C does not exist because of early expiration of the first node. The network lifetime is affected due to the fact that nodes are available as relays for a longer time, hence not enough battery charge is left below threshold for use as a source or destination, and therefore nodes becomes exhausted.

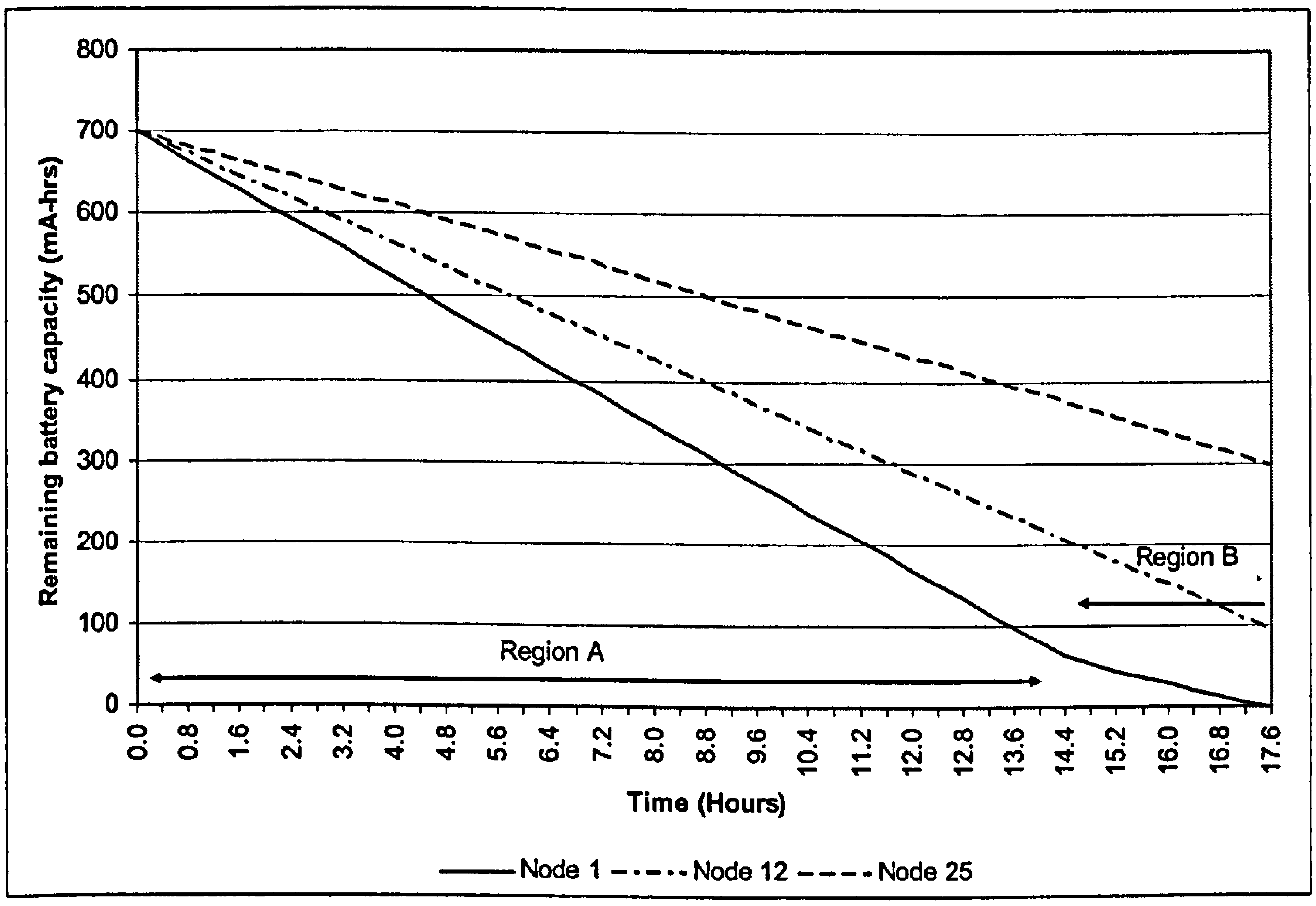


Figure 4.3 Battery charge vs time (lower threshold – 10% of maximum battery charge)

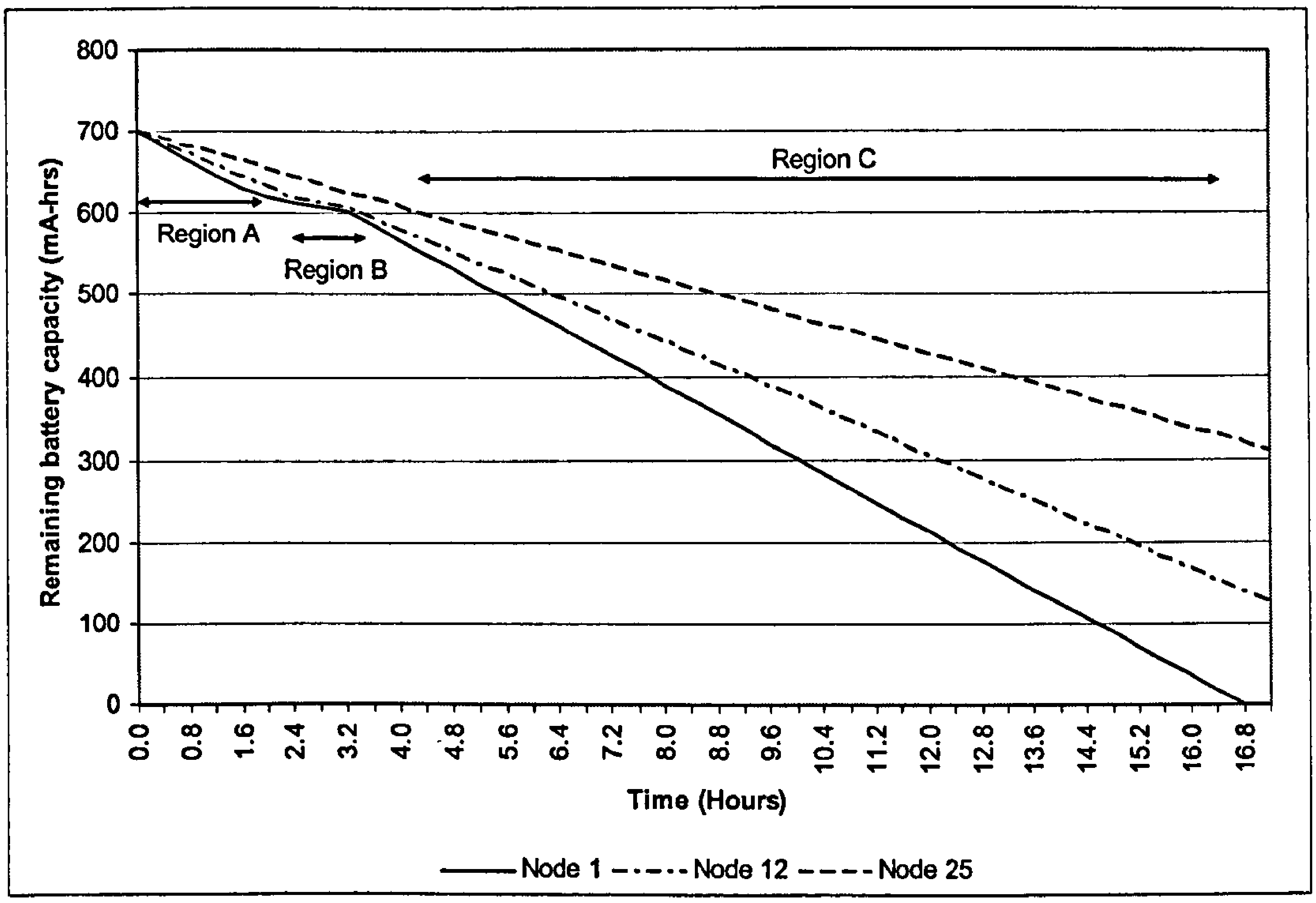


Figure 4.4 Battery charge vs time (higher threshold – 90% of maximum battery charge)

- **Using higher thresholds**

Figure 4.4 shows the remaining battery against time with a threshold set at 90% of the maximum battery charge. When a higher battery charge threshold is selected, nodes are isolated from relays at an early stage. When all nodes reach threshold, they are restored as relays. In this case, the period of region B is small because the disparity between the most used nodes and the least used nodes when the most used node reaches threshold is very small. The improvement in the network lifetime compared to the case with MPR is therefore very small.

- **Optimum threshold**

Figure 4.5 shows the remaining battery against time with a threshold set at 25% of the maximum battery charge. It can be seen that the most heavily node reaches threshold first and then its usage is reduced to functioning only as a source node and other nodes are used to form the required routes. The minimum power route is calculated using the available nodes. As the number of nodes reaching threshold increases, the current required for transmission by the remaining nodes increases. The battery charge of the more heavily used nodes is reduced until all nodes have reached the threshold. At that point, all nodes are again made available for use as a relay and the most heavily used node is the first to reach exhaustion. At the optimum threshold, the network lifetime is about 23 hours which is about 44% improvement over the case with MPR.

The optimum threshold is the value that leads to all nodes reaching zero battery charge at the same time. Further investigation has been carried out to evaluate network lifetime with different battery charge thresholds. Figure 4.6 shows that the optimum battery charge threshold is 25%.

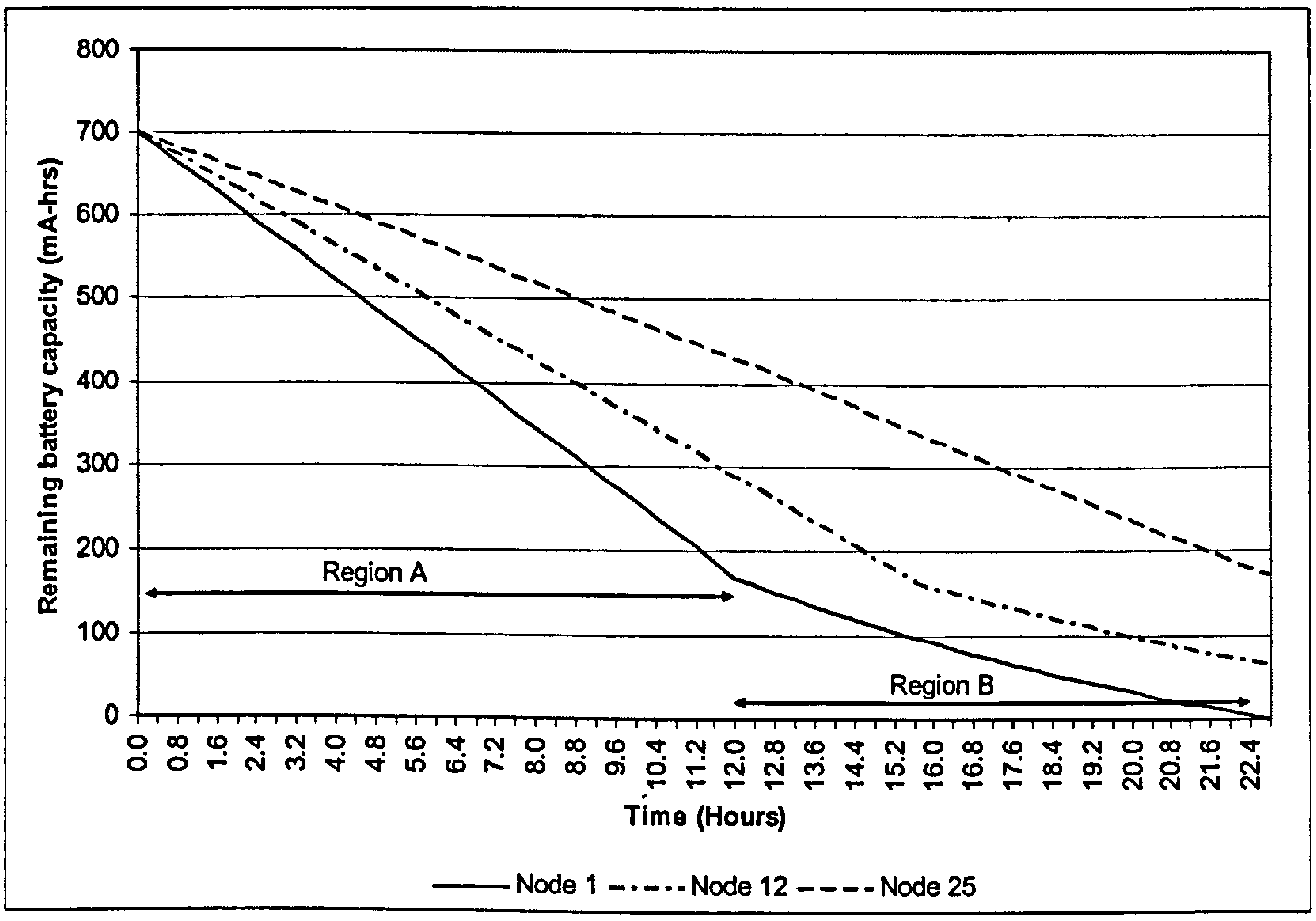


Figure 4.5 Battery charge vs time (optimum threshold – 25% of maximum battery charge)

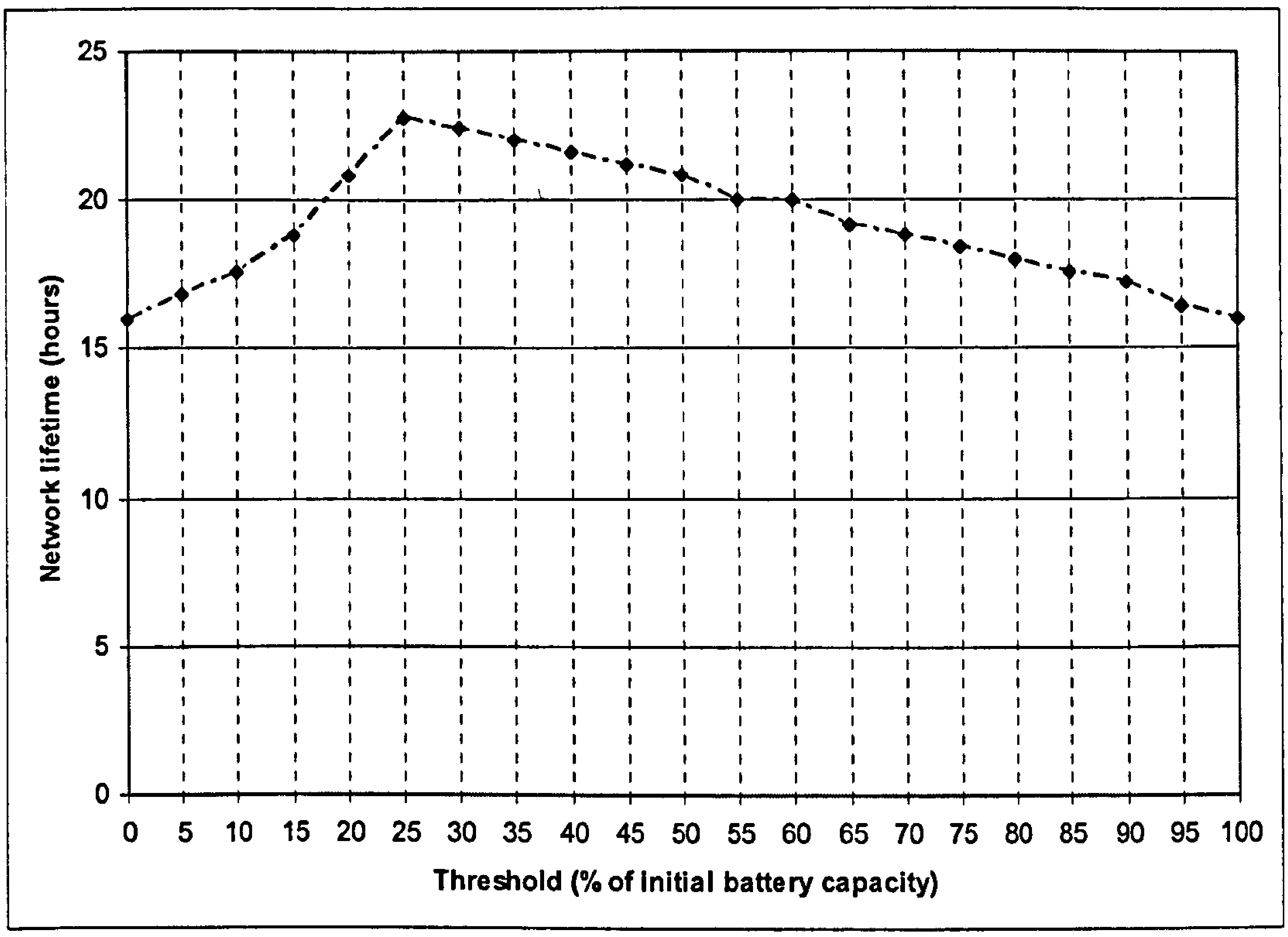


Figure 4.6 Network lifetime vs battery charge thresholds

4.5 Analysis of residual battery charge routing scheme

In the previous schemes described in sections 4.3 and 4.4, transmission power is an important routing metric so that minimum power route can be selected. However continuous use of the minimum power route via a specific relay node will deplete node's battery charge rapidly. Such problem can be solved if battery charge threshold is used (as discussed already in the section 4.4). Another way is to incorporate the remaining battery charge as a routing metric cost [105]. Suppose a node's willingness to forward packets is a function of its remaining battery charge. As the battery charge decreases, the value of the metric cost for the node increases. The cost (R_j) for the route is calculated using the following equation:

$$R_j = \min \sum_{i=0}^{D-1} \left(\frac{1}{B_i} \right) \tag{4.12}$$

where

B_i Residual battery charge of node i (mAhrs)

D Total number of nodes in the route; D-1 is the number of relay nodes in the route

The above equation avoids routes which include nodes with a low battery charge. Figure 4.7 illustrates this process. In this network the source and destination nodes are not connected directly due to clutter in between them, and there are two routes between them, route A and route B. Route A is the lowest power route, but it includes a relay node with the lowest remaining battery charge.

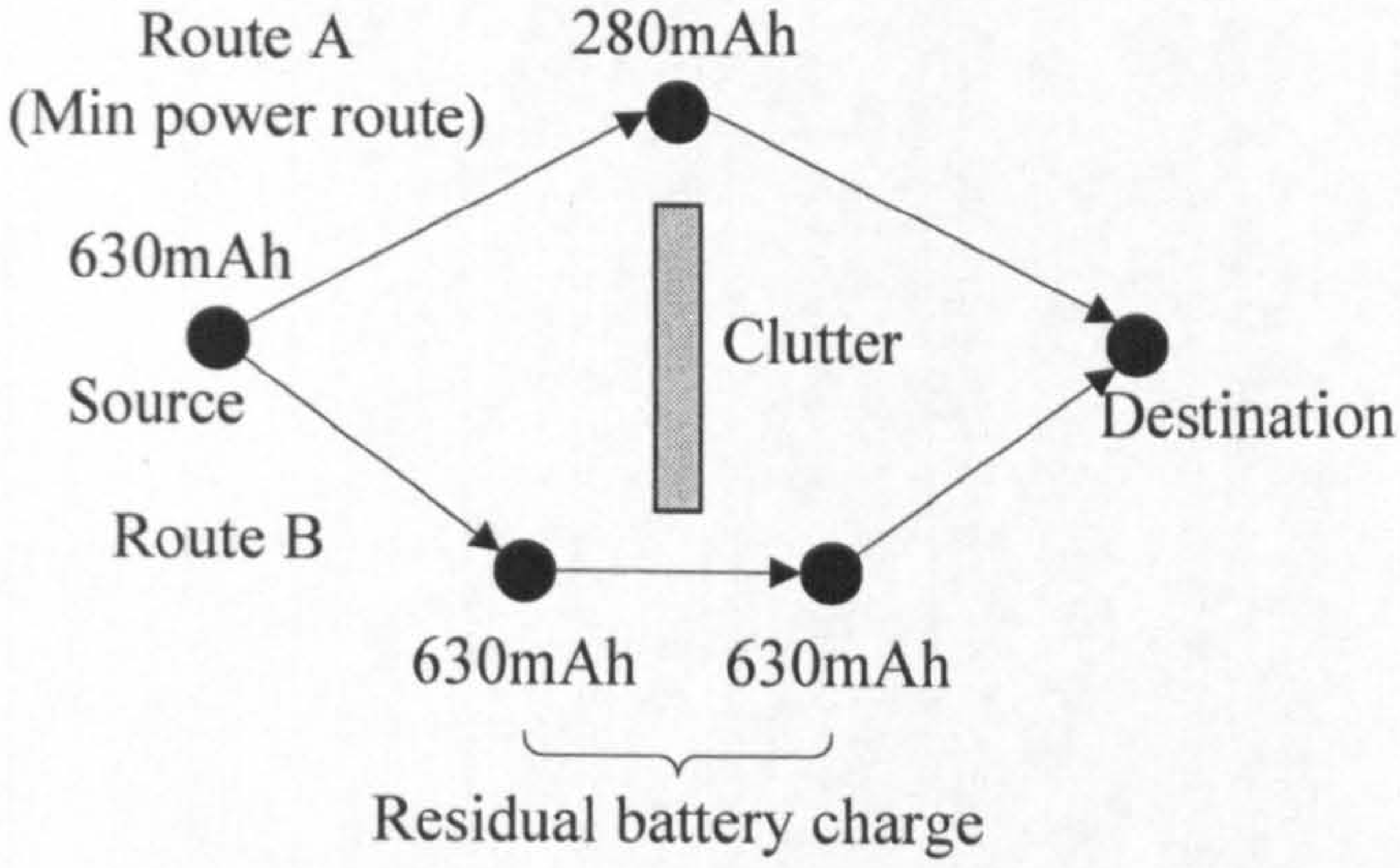


Figure 4.7 Residual battery charge scheme

If equation 4.12 is used to calculate the route then the calculated cost for Route A and Route B are:

R_j (route A): (1/630) + (1/280) = 0.00515

R_j (route B): (1/630) + (1/630) + (1/630) = 0.00476

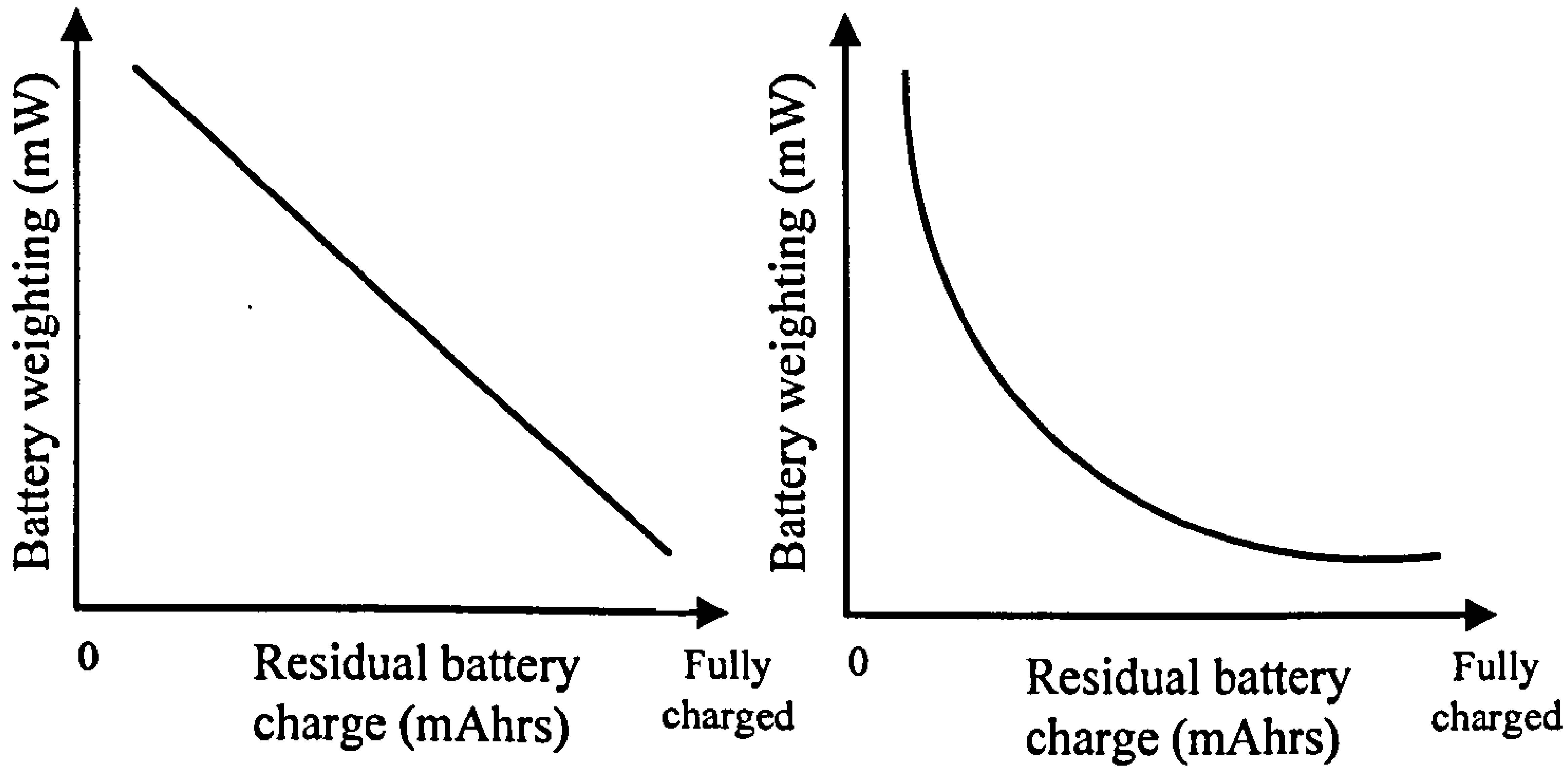
Route B is calculated to have the lowest cost on this basis and so the node with the low battery charge is avoided.

The advantage of this scheme is that low battery charge nodes are avoided. However, the alternate route may contain several hops which increase the total end to end power consumption.

4.6 Minimum Power Routing/Maximum Battery Lifetime (MPR/MBL) scheme

In this scheme, the routing takes into account of the minimum power and residual battery charge. Minimum power routing is used when all the nodes have full charge batteries. However, in order to prevent nodes from being overused, residual battery charge is translated into battery weights (in terms of power), so that when a node's battery charge decreases, the battery weight increases.

The relationship between the battery weighting and the residual battery charge can be selected to optimise the balance between battery usage and minimising the power consumption. Two characteristics have been analysed, a linear relationship and an inverse relationship. Figure 4.8(a) shows that the battery weight increases linearly when the node's residual battery charge decreases. Figure 4.8(b) shows that the battery weight increases inversely when the node's residual battery charge decreases.



A) Linear relationship

A) Inverse relationship

Figure 4.8 Linear and inverse relationships between residual battery charge and battery weighting

4.6.1 MPR/MBL with linear battery weighting

$$R_L = \min \sum_{i=0}^{j-1} (P_i + W_i) \tag{4.13}$$

where

R_L = cost of the route between the source and the destination node

i = (0, 1, 2, 3, ...j-1, j); 0 = source node, 1, 2, 3 are relay nodes, j-1 is the last relay node in the route, and j is the destination node

P_i = Average power which includes transmission and processing powers (mW)

W_i = battery weighting of node i (translated into mW)

Equation 4.13 consists of two terms; the first term is related to the power. This power is the average power consumption per node which includes transmission and processing powers. The second term is related to the residual battery charge. When all the nodes have fully charged batteries, then the battery weighting term is negligible and the route selection depends on the power term, and the network behaves the same as in the case of MPR (discussed in section 4.4.1). However, when the battery charge is low, the battery weighting becomes the dominant term and the route selection discourages nodes with low battery charge.

For example, a node selects the value of battery weighting such that it equals the average power when the battery charge reduces to 200mAhrs. The battery charge at which the battery weighting term is equal to the average power term is called cross-over point. This is shown in figure 4.9.

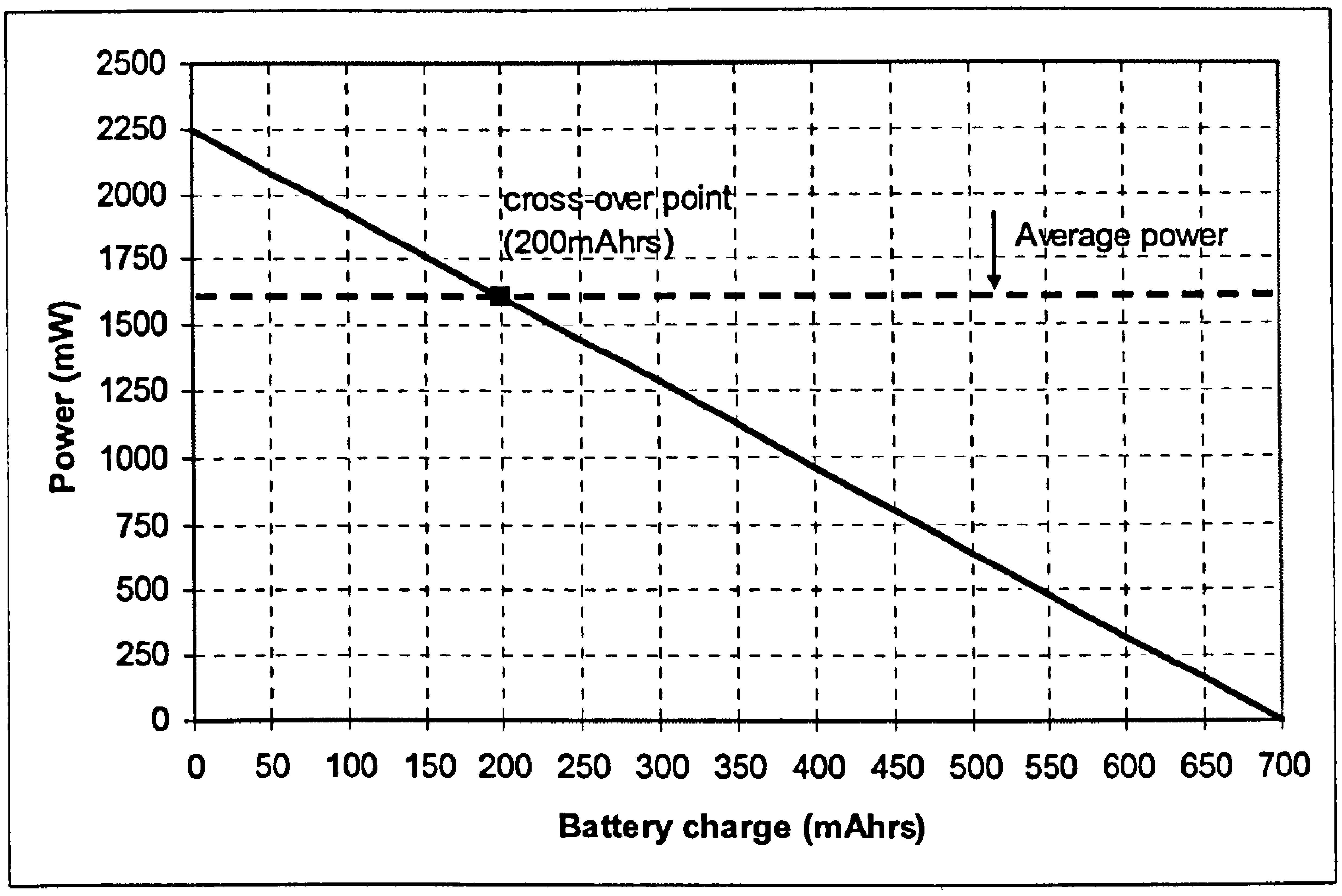


Figure 4.9 Cross-over point at 200mAhrs

The linear weighted battery charge is calculated according to the following equation:

$$y = mx + c \tag{4.14}$$

where

y = average power which includes transmission and processing powers (mW)
m = slope (mW/mAhrs)
x = residual battery charge (mAhrs)
c = constant (mW)

When the nodes have fully charged batteries (x_1=700mAhrs), the corresponding battery weight is zero (y_1=0mW). Similarly when the nodes have battery charge left about 200mAhrs (x_2=200mAhrs), the corresponding battery weight is (y_2=1606mW). The slope of the line 'm' is $(y_2-y_1)/(x_2-x_1)$ which gives -3.21mW/mAhrs.

MPR/MBL with linear battery weighting is investigated using a example network shown in figure 4.10.

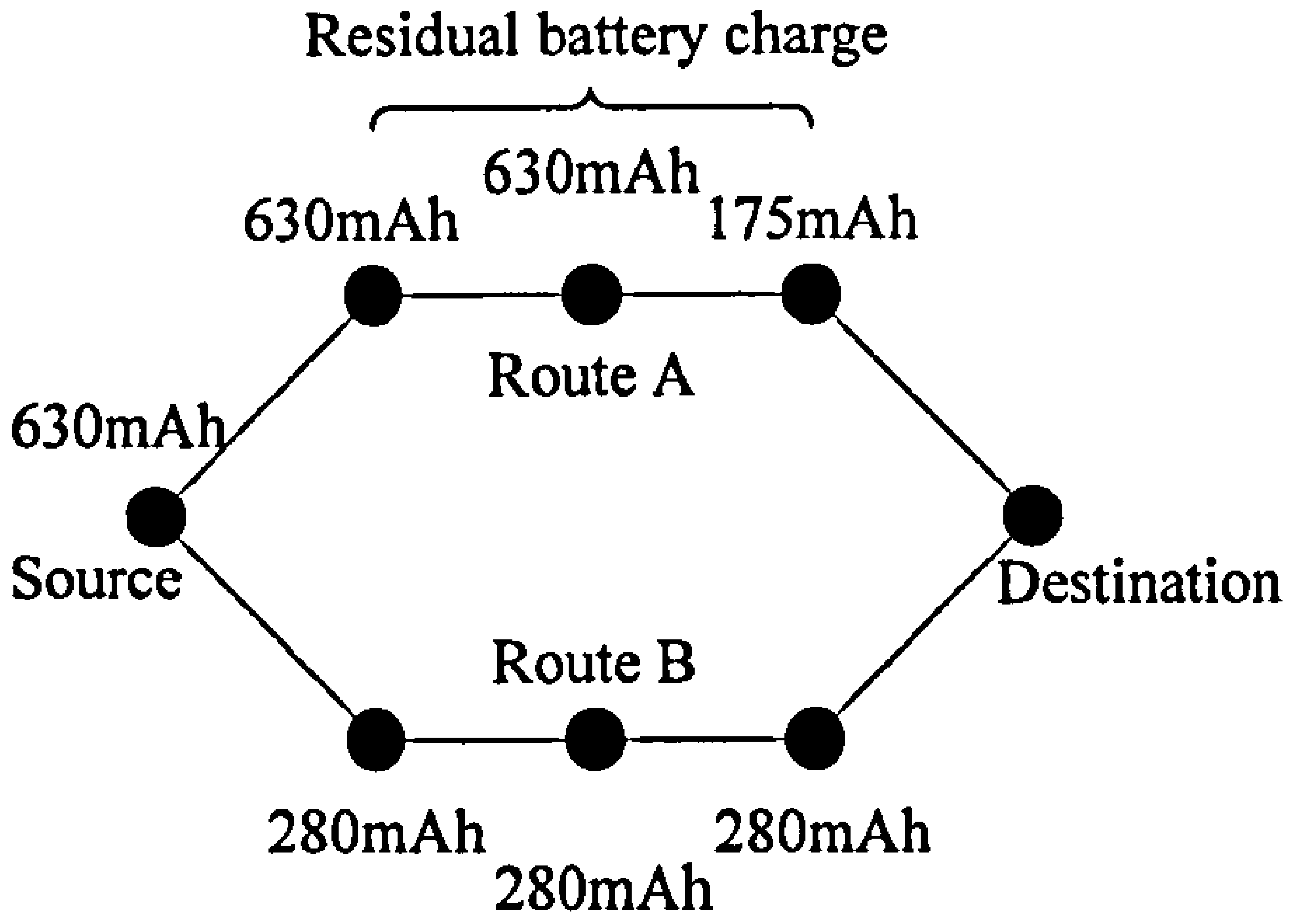


Figure 4.10 Example Network

The average signal processing power per node due to transmission is 722mW (219mAx3.3v).

The average signal processing power per node due to reception is 710mW (215mAx3.3v).

The total average signal processing power per node is 1432mW (722mW+710mW).

The average transmission power per node is 174mW.

The average power consumption per node is 1432mW + 174mW = 1606mW.

Using figure 4.9, the battery charge weighted values for the nodes in Route A are:
630mAh node = 255mW
175mAh node = 1686mW

The battery charge weighted value for the nodes in Route B are:
630mAh node = 255mW
280mAh node = 1349mW

Hence, using equation 4.13, the route calculation gives:

Cost of route A:

[(1606mW + 255mW) x 3] + (1606mW + 1686mW) = 8875mW

Cost of route B:

(1606mW + 255mW) + [(1606mW + 1349mW) x 3] = 10726mW

Route A has the minimum cost between the source and the destination node.

Further analysis is carried out in which a node selects the value of battery weighting such that it equals the average power when the battery charge reduces to 600mAhrs.

This increases the battery charge weighting with the aim of giving more protection to nodes with a low battery charge.

Figure 4.11 shows the cross-over point which correspond to the residual battery charge of 600mAhrs and an average power of 1606mW.

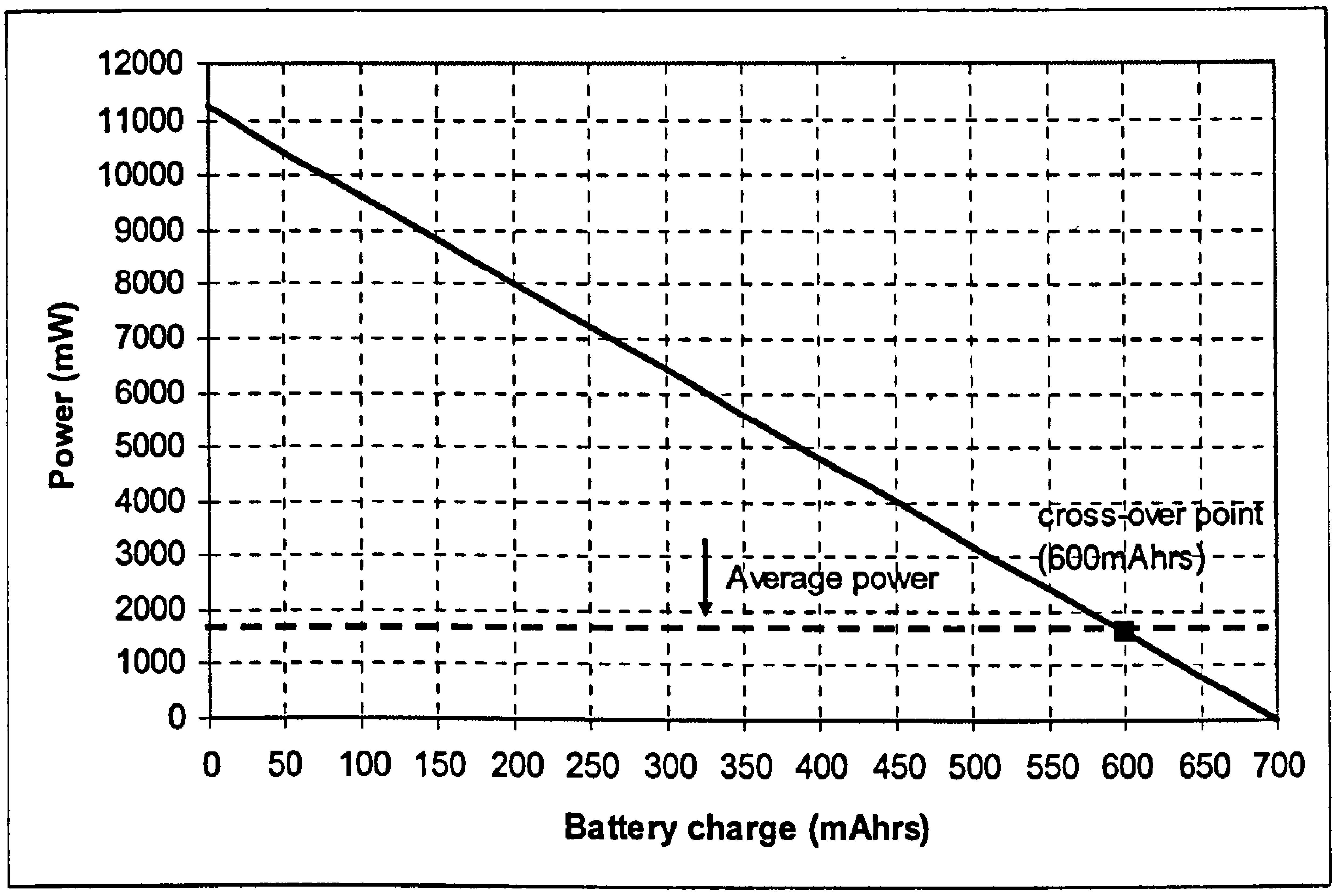


Figure 4.11 Cross-over point at 600mAhrs

In this case, using figure 4.11, the battery charge weighted values for the nodes in Route A are:

630mAh node = 1124mW

175mAh node = 8430mW

The battery charge weighted values for the nodes in Route B are:

630mAh node = 1124mW

280mAh node = 6744mW

Hence, using equation 4.13, the route calculation gives:

Cost of route A:

[(1606mW + 1124mW) x 3] + (1606mW + 8430mW) = 18226mW

Cost of route B:

(1606mW + 1124mW) + [(1606mW + 6744mW) x 3] = 27780mW

Again Route A has the minimum cost between the source and the destination node.

4.6.2 MPR/MBL with inverse battery weighting

In this scheme, the battery weight increases sharply when the residual battery charge reaches over the cross-over point. The cost (R_E) is calculated according to the following equation:

$$R_E = \min \sum_{i=0}^{j-1} \left(P_i + \left(\frac{C}{x_i} \right)^a \right) \tag{4.15}$$

where,

R_E = cost of the route between the source and the destination node

$i = (0, 1, 2, 3, \ldots j\text{-}1, j)$; 0 = source node, 1, 2, 3 are relay nodes, j-1 is the last relay node in the route, and j is the destination node

P_i = Average power which includes transmission and processing powers (mW)

C = constant (mAhrs.mW)

x_i = residual battery charge (mAhrs)

a = exponent

In equation 4.15, the first term calculates the end to end power, and the second term calculates the residual battery charge weighting. The difference in the second term in this equation as compared to the equation 4.13 is that the battery weight increases inversely when node's battery is used.

The cross over point can be set so that when the available battery of the node reaches the cross over point, the battery weighting term becomes larger than the average transmission power term and this weights against using this node in the selected route. A value is selected for constant 'C' which gives an average power of 1606mW when the residual battery charge is 200mAhrs. Figure 4.12 shows the relationship between the battery weighting and the residual battery charge for different values of 'a' (exponent orders). Figure 4.12 shows that the average power term is larger than the battery weighting term until the battery charge reduces to 200mAhrs. The battery weight increases rapidly as the battery charge reduces further. The rate of increase depends on the value of the exponent 'a'.

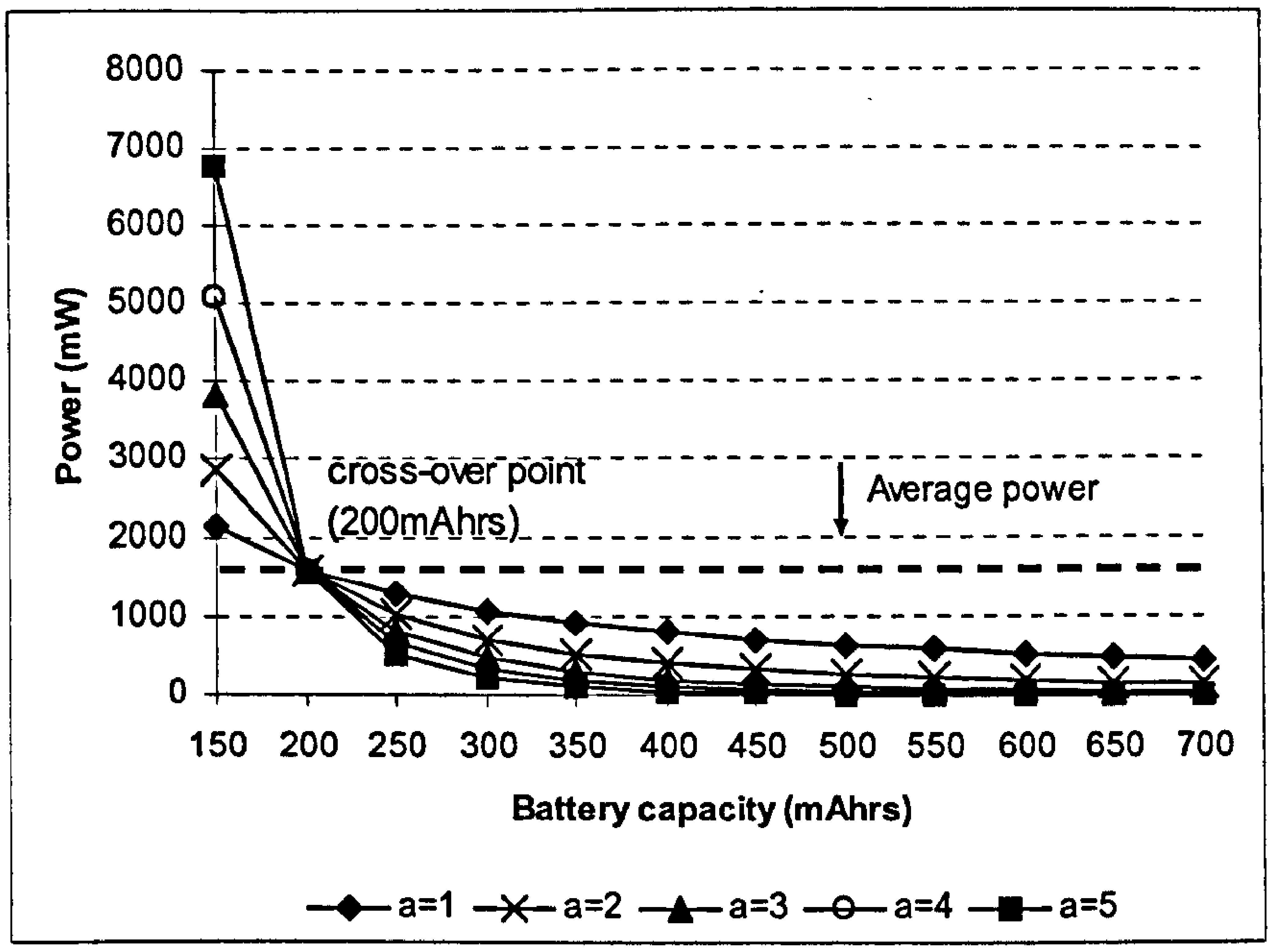


Figure 4.12 Inverse battery weighting (cross over point 200mAhrs)

MPR/MBL scheme with inverse battery weighting (cross-over point of 200mAhrs) is analysed using the example network shown in figure 4.12.

Using figure 4.12 (with exponent order a=1), the battery charge weighted values for the nodes in Route A are:
630mAh node = 510mW
175mAh node = 1835mW
The battery charge weighted values for the nodes in Route B are:
630mAh node = 510mW
280mAh node = 1147mW
Hence, using equation 4.15, the route calculation gives:
Cost of route A:
[(1606mW + 510mW) x 3] + (1606mW + 1835mW) = 9789mW
Cost of route B:
(1606mW + 510mW) + [(1606mW + 1147mW) x 3] = 10375mW

Route A will be selected because the total end to end power is lower than route B. As a result, the relay node with low available battery charge will reach zero battery charge at an early stage.

When a higher value of the exponent 'a' is used, the battery weighting increases more rapidly as the residual battery charge reduces below the cross-over point. Using the same example as before with an exponent order a=2 for the calculation below:

Using figure 4.12, the battery charge weighted values for the nodes in Route A are:
630mAh node = 162mW
175mAh node = 2098mW
The battery charge weighted values for the nodes in Route B are:
630mAh node = 162mW
280mAh node = 819mW
Hence, using equation 4.15, the route calculation gives:
Cost of route A:
[(1606mW + 162mW) x 3] + (1606mW + 2098mW) = 9008mW
Cost of route B:
(1606mW + 162mW) + [(1606mW + 819mW) x 3] = 9043mW

Route A has the minimum cost between the source and the destination node.
Using figure 4.12 (with exponent order a=3), the battery charge weighted values for the nodes in Route A are:
630mAh node = 51mW
175mAh node = 2397mW
The battery charge weighted values for the nodes in Route B are:
630mAh node = 51mW
280mAh node = 585mW
Hence, using equation 4.15, the route calculation gives:
Cost of route A:
[(1606mW + 51mW) x 3] + (1606mW + 2397mW) = 8974mW
Cost of route B:
(1606mW + 51mW) + [(1606mW + 585mW) x 3] = 8230mW

At exponent order (a=3), route B is the lowest cost route. It is clearly seen in figure 4.12 that at the cross over point, the higher order exponent curve rises sharply, hence routes including nodes with a low residual battery charge are avoided.

A cross-over point of 600mAhrs is also analysed using the example network shown in figure 4.10. In this case the value of constant 'C' is chosen such that the average power is equal to 1606mW for a residual battery charge of 600mAhrs. Figure 4.13 shows that the average power term is larger than the battery weighting term until the battery charge reduces to 600mAhrs. The battery weight increases rapidly as the battery charge reduces further.

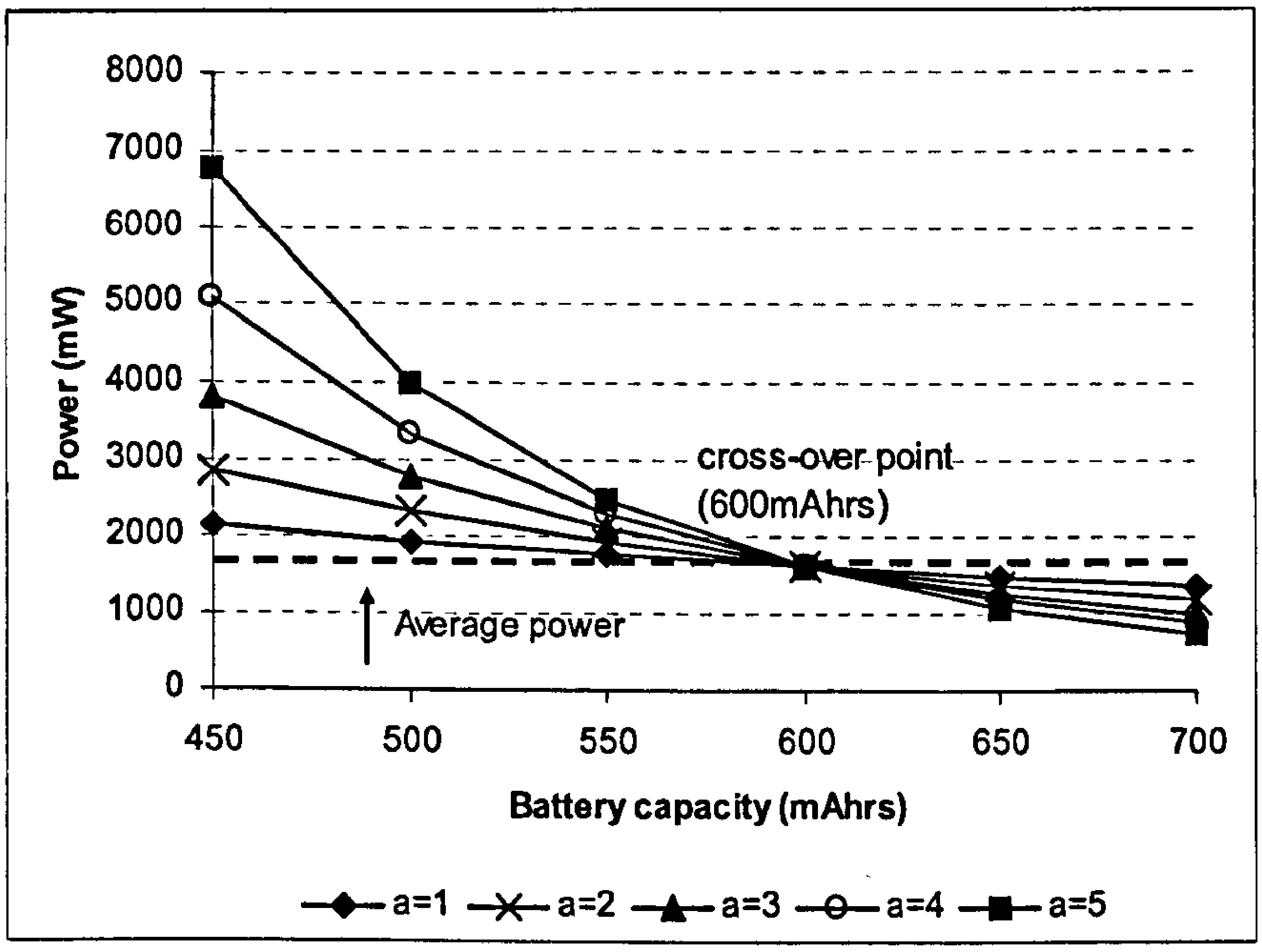


Figure 4.13 Inverse battery weighting (cross over point 600mAhrs)

Using figure 4.13 (with exponent order a=1), the battery charge weighted values for the nodes in Route A are:

630mAh node = 1530mW

175mAh node = 5506mW

The battery charge weighted values for the nodes in Route B are:

630mAh node = 1530mW

280mAh node = 3441mW

Hence, using equation 4.15, the route calculation gives:

Cost of route A:

[(1606mW + 1530mW) x 3] + (1606mW + 5506mW) = 16520mW

Cost of route B:

(1606mW + 1530mW) + [(1606mW + 3441mW) x 3] = 18277mW

Using figure 4.13 (with exponent order a=2), the battery charge weighted values for the nodes in Route A are:
630mAh node = 1457mW
175mAh node = 18879mW
The battery charge weighted values for the nodes in Route B are:
630mAh node = 1457mW
280mAh node = 7374mW
Hence, using equation 4.15, the route calculation gives:
Cost of route A:
[(1606mW + 1457mW) x 3] + (1606mW + 18879mW) = 29674mW
Cost of route B:
(1606mW + 1457mW) + [(1606mW + 7374mW) x 3] = 30003mW

At exponent orders (a=1 and a=2), Route A has the minimum cost between the source and the destination node.

Using figure 4.13 (with exponent order a=3), the battery charge weighted values for the nodes in Route A are:
630mAh node = 1387mW
175mAh node = 64727mW
The battery charge weighted values for the nodes in Route B are:
630mAh node = 1387mW
280mAh node = 15802mW
Hence, using equation 4.15, the route calculation gives:
Cost of route A:
[(1606mW + 1387mW) x 3] + (1606mW + 64727mW) = 75312mW
Cost of route B:
(1606mW + 1387mW) + [(1606mW + 15802mW) x 3] = 55217mW

In both the cases (cross-over points of 200mAhrs and 600mAhrs), when a =3, route A is avoided and the node with the low battery charge is not used.

The inverse relationship described in equation 4.15 allows independent control of the power and residual battery terms. This allows the network lifetime to be optimised. The ability to be able to separately control the power and the residual battery charge can be used to optimise the route selection and maximise the network lifetime.

4.6.3 Comparison of MPR/MBL with power aware routing scheme

Power aware routing discussed in section 4.3.2 is analysed using the example network shown in figure 4.10. At low exponent order ($\alpha=1$) in equation 4.4, the link cost of route A is:

[174mW x (700mAhrs/630mAhrs)] x 3 + [174mW x (700mAhrs/175mAhrs)] = 1276mW

Link cost of route B is:

[174mW x (700mAhrs/630mAhrs)] + [174mW x (700mAhrs/280mAhrs)] x 3 = 1498mW

Route A has the minimum cost between the source and the destination node.

Using equation 4.4 ($\alpha=2$), the link cost of route A is:

[174mW x (700mAhrs/630mAhrs)^2] x 3 + [174mW x (700mAhrs/175mAhrs)^2] = 3428mW

Link cost of route B is:

[174mW x (700mAhrs/630mAhrs)^2] + [174mW x (700mAhrs/280mAhrs)^2] x 3 = 3477mW

Route A has the minimum cost between the source and the destination node.

Using equation 4.4 ($\alpha=3$), the link cost of route A is:

[174mW x (700mAhrs/630mAhrs)^3] x 3 + [174mW x (700mAhrs/175mAhrs)^3] = 11852mW

Link cost of route B is:

[174mW x (700mAhrs/630mAhrs)^3] + [174mW x (700mAhrs/280mAhrs)^3] x 3 = 8395mW

When a higher exponent order is used ($\alpha=3$), then the link cost of route A increases mainly due to low battery charge node (175mAhr) in the route. Route B has the minimum cost between the source and the destination node. As compared to

MPR/MBL, power aware routing lacks separate control of the transmission power and the residual battery terms. This is further discussed using a simulation model in section 6.10, chapter 6.

4.7 Summary

Various reported schemes have been analyzed for maximising the lifetime of an ad hoc network. In multi hop ad hoc networks where the terminals are battery powered, if minimum power routing is used, then nodes in a central location will tend to be used more than nodes at the periphery and this reduces the lifetime of the network. To enable the power consumption to be evenly spread between all the nodes in the network, a threshold can be set for the remaining battery charge, such that when a battery reaches this threshold, the node is withdrawn from use as a relay. The threshold scheme described by Toh [15] is analyzed to find the optimum threshold and the improvement in the network lifetime that can be achieved, with the time to first node exhaustion being the critical value. A theoretical 25 node model has been developed, and the effect of setting various battery charge thresholds has been investigated. The results show that the network lifetime is dependent on the threshold; at the optimum threshold of 25%, the network lifetime is improved by 44% over MPR.

A routing scheme based on residual battery charge is analyzed in which the route with the largest residual battery charge is selected. However the selected routes with this scheme are not necessarily the minimum power routes.

In the proposed MPR/MBL scheme, minimum power and residual battery charge schemes are combined in order to optimise the balance between the minimum power routes and the maximum network lifetime. This is achieved using a combination of power consumption and residual battery charge values. Linear and inverse characteristics have been investigated and the inverse form is shown to be preferable because battery weights increases inversely when the node's residual battery charge decrease, hence nodes are protected from being overused. This allows the network lifetime to be optimised.

5. System Design

5.1 Introduction

This chapter describes the design of an ad hoc network in which nodes can function as relays to form multi hop routes across the network. The nodes in an ad hoc network carry out all the signal processing functions required to support routing and access management in addition to the data transmission. In many case the terminals in these networks are battery operated to allow mobility. Power minimisation is therefore one of the key design objectives in an ad hoc network in order to maximise the network lifetime.

This chapter describes how schemes for maximising the network lifetime can be implemented in a network. The chapter describes the network design and the network signalling.

5.2 Network design assumptions

The requirements and assumptions for the system design are as follows:

a) An ad hoc network is considered which is suitable for use by emergency services where there is no communication by fixed infrastructure.
b) The nodes are powered and they cannot be recharged during the network operation.
c) The network size is 50m x 50m (determined by the power budget).
d) The number of nodes in the network is 25 (compatible with the size of the network).
e) The speed of the nodes is in the range 0-100 km/hr.

5.3 Network design requirements

5.3.1 Physical layer

The physical layer design is based on the IEEE 802.11 specification [12] in order to be compatible with existing components. The radio frequency is 2.4GHz and the maximum transmission power is +20dBm. The RF power amplifier efficiency is 40% [102]. The design assumes the use of a single frequency for an end to end path and each node uses an omni-directional antenna. The target bit rate is 6 Mbps to allow voice, data and video applications.

5.3.2 Medium access layer

Medium access is based on a busy tone scheme as described in section 2.2.1.4, chapter 2. Each node indicates the received power and shares this information with other nodes in the network. The received power information is distributed in the network with the help of an out of band control channel, details of which are given in the next section. With the knowledge of the received power, each node is aware of on-going communication in the neighbourhood, calculates the minimum allowed transmit power, and hence interference can be reduced.

5.3.3 Network layer

The distance vector routing protocol sends routing table to its directly connected neighbour. In the link state protocol, each node sends link state changes to all other nodes in the network. Link state routing protocols generate larger routing control overhead than distance vector. In large networks with mobile nodes, the transmission of routing update will consume most of the bandwidth which makes link state routing unfeasible for bandwidth limited wireless ad hoc networks or if nodes are powered by limited battery. In order to conserve bandwidth and power, distance vector routing is used in the system design.

The routing decision is made at the source node. For example, if minimum power routing is used, then the source node determines all the possible routes to the destination node and selects the minimum power route. Similarly, if residual battery charge scheme is used, then the source node determines which relay nodes have higher residual battery charge and then sends data through these relay nodes. The data packet originates from the source node contains all the relay node IDs to be traversed to the destination node. The relay nodes forward data packet to the next relay node without searching for the route. This technique is called source routing.

As discussed in Chapter 2, there are two classes of routing protocol, proactive and reactive and the best approach depends on the speed of the nodes and the frequency at which traffic is generated by nodes. To accommodate all conditions, a hybrid approach is selected here that can select either a proactive or a reactive approach depending on the prevailing network conditions. This is further discussed in the section 5.4.1.1.

5.3.4 Transport layer

Modified TCP is used in which congestion and packet loss is differentiated with the help of ECN and ELN mechanisms respectively.

5.4 Network signalling

One of the key features of this design approach is to enable nodes that cannot communicate directly, to be able to connect through relay nodes, because routing depends on information from all of the other nodes in the network. The path loss, available battery and allowed transmission power are required to carry out route calculations. Such information is shared with the help of signalling. The signalling can be done in one of two ways; in-band or out-of-band. The in-band approach uses the same channel for control and data, the out-of-band approach uses a separate channel.

For example, IEEE 802.11 uses in-band signalling. When a connection is setup, one channel is allocated which is used for both data and network signalling. Network signalling consists of management and control frames. One of the management frames is a beacon frame which broadcasts the existence of a network and other necessary information required by the station to join the network such as a timestamp, the beacon interval, and the maximum allowed transmission power set by the regulatory body of the country [12]. RTS/CTS signalling can be used to reserve nodes for data transmission.

In an ad hoc network, each node can share information with other nodes in the network with the help of beacon frames. Such information is useful to perform routing. If an in-band approach is used in which the same channel is used for data and control signals, then in the presence of higher network traffic, beacon frames may not be transmitted in the network. As a result, other nodes in the network will not have up to date information which may obsolete previously gathered route.

Out-of-band signalling is proposed in which beacon frames are transmitted on a different frequency channel than the data. In addition, contention in the control channel for beacon frame transmissions is managed by allocating a time slot to each node.

5.4.1.1 Beacon frame format

The beacon frame carries information about other nodes in the network including the attenuation between the links, the available battery charge of the nodes, and the maximum allowed transmission power. The beacon frame is broadcasted in the network.

The beacon frame contains (4+2*N*) bytes for each node in the network, where *N* is the number of nodes in the network. For a network with *N* nodes, there will be *N* time 'slots' available in each 'full frame'. Each node is transmitting a full frame of size *N*(4+2*N*) bytes. Figure 5.1 shows the beacon frame format and the allocation of time slots.

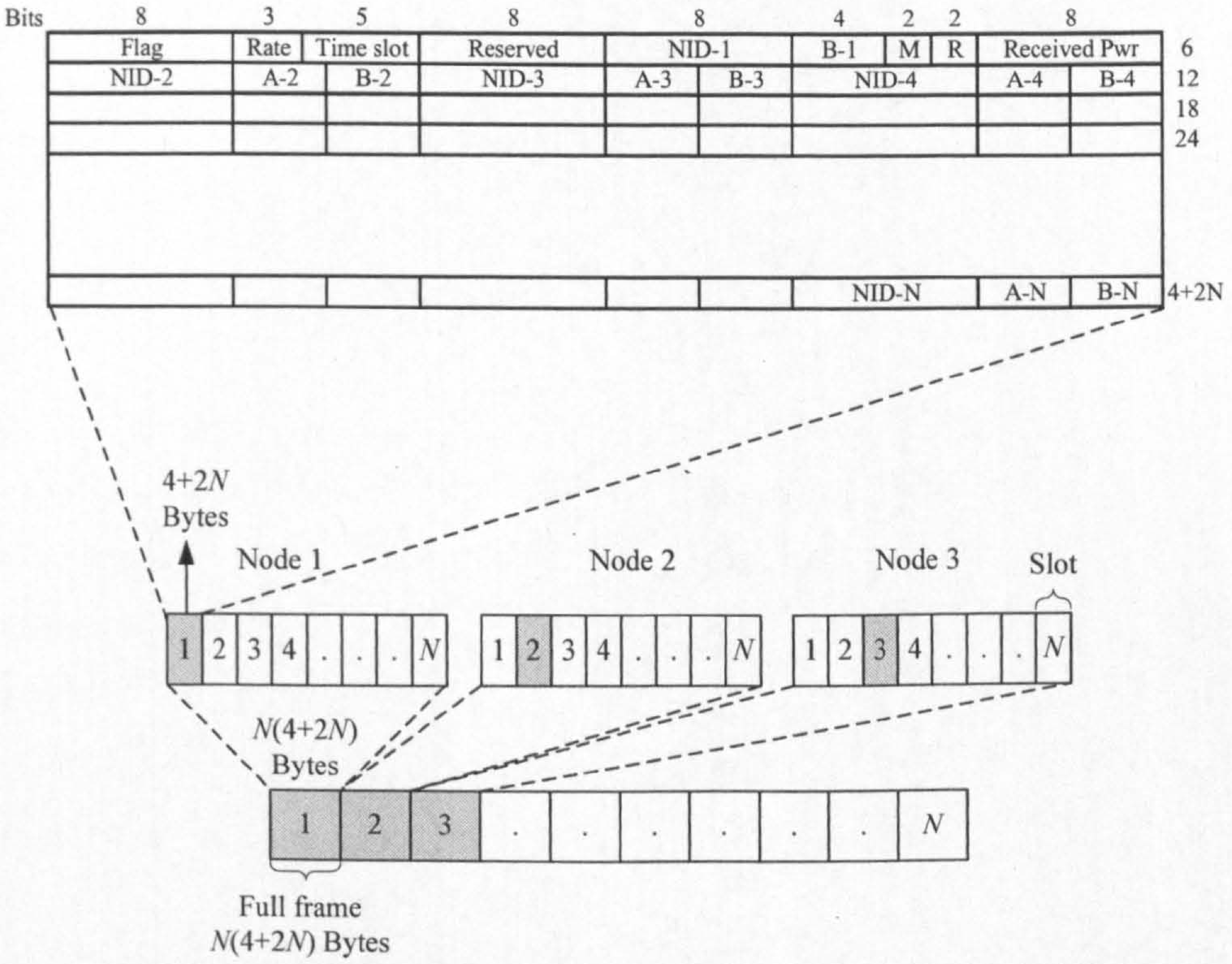


Figure 5.1 Beacon frame format and time slot allocation

The “Flag” field is used for frame alignment. In the beginning of each beacon frame, a unique pattern 011111110 is used. An ending flag is not necessary in the beacon frame because a fixed beacon frame size is used in the system design. Receivers are continuously monitoring this flag sequence to align on the start of a frame. If the same pattern appears inside of the beacon frame (except flag field), then in order to prevent frame synchronization, bit stuffing is used in which the transmitter adds an extra 0 bit after each occurrence of five 1s in the frame. A similar frame alignment process is used in High level Data link Control (HDLC) protocol [106].

The “Rate” field is used to select beacon channel bit rate which is discussed further in section 5.4.1.4.

In the system design, it is assumed that a pre-defined time slot is allocated to each node. However, dynamic time slot assignment can be used which requires an additional field in the beacon frame to process the time slot requirement. Further work is needed in this area to investigate dynamic time slot allocation process in an ad hoc network without using a dedicated system to manage time slots between the nodes.

The “Reserved” and “R” fields are reserved for future use.

The “NID-1” is the node ID of the originating beacon frame (node 1).

The “B-1” is the battery status of node 1.

The “M” determines the resource discovery mode. This mode can be configured to be continuous or on-demand. If link characteristics are changing continuously, then it is beneficial to use continuous resource discovery. This stores accurate link information in the database and establishes the route without adding delay. However, if the network is in a state where the link characteristics do not change, then the nodes can use on-demand resource discovery to reduce power consumption. For this reason, the beacon frame contains a mode field for setting a continuous or on-demand mode.

Link characteristics can be determined by calculating the attenuation between the nodes with the help of beacon frames. The attenuation may be varying significantly due to the movement of the nodes or movement of objects in the transmission path. Nodes indicate this link characteristic by selecting Bit A = 1 and will proceed to transmit in continuous mode. If any node receives a beacon signal indicating another node is using the continuous mode using Bit A = 1 then it will also switch to continuous mode to co-operate. If all nodes indicate that the on-demand mode is acceptable by setting Bit A = 0, then all nodes can switch to the on-demand mode. If all nodes are in on-demand mode and a node wants to find the route, then Bit B = 1 is used to trigger a route discovery.

When a node first joins the network, it does not know which mode is being used by other nodes so it first listens to the on going beacon frames and adapts its beacon frame transmission according to the Bit A. If continuous operation is in use, then the newly joined node can use a continuous beacon transmission mode. However, during the beacon frame transmission, each node gets attenuation information from other nodes. If the attenuation does not vary significantly then the node can select on demand mode (Bit A = 0). When any node broadcasts a route discovery frame when operating in an on-demand mode, it sets Bit B = 1 then other nodes in the network respond to this route discovery by sending beacon frames. When data transmission is completed to the destination node, then this route is saved in the cache for future use.

	Bit A	Bit B
Operating mode - continuous	1	
Operating mode - on demand	0	
Trigger resource discovery		1
Triger inactive		0

Table 5.1 Mode field bit definitions

The "Receiver Pwr" field contains the received power information which is used to calculate the allowed power by the transmitter node in order to prevent interference at the receiver node. Simultaneous communication is described in the enhanced system design in chapter 7. In chapter 5, simultaneous communication is not assumed, so this field is not used.

Fields from the second row of the beacon frame in figure 5.1 contain information about the neighbouring nodes (directly connected nodes). "NID-2" is the ID of the neighbouring node 2. The "A-2" field contains the attenuation between node 1 and node

2. Beacon frames are transmitted at fixed power, so by measuring the received power of the beacon frame, the attenuation between each node can be calculated. The "B-2" field gives the status of the node 2 battery.

5.4.1.2 Path loss model

A two-ray ground reflection path loss model [107] is used to calculate the distance between the nodes, this is given by:

$$PL(dB) = 40\log(d) - 10\log(G_t) - 10\log(G_r) - 20\log(H_t) - 20\log(H_r) \tag{5.1}$$

where

d = distance between the source and destination in meters.

G_t and G_r = Gain of the transmitter and receiver respectively; assumed omni directional antenna with a unit gain.

H_t and H_r = Height of the transmitter and receiver in meters respectively; the average user height is assumed to be 1.5m.

Equation 5.1 is further simplified by adding G_t and G_r antenna gain = 1, and H_t and H_r = 1.5m:

$$PL(dB) = 40\log(d) - 7.04 \tag{5.2}$$

In addition, a random clutter is included which is independent of the distance between the nodes. This models the attenuation that would be experienced due to buildings and trees in a typical application environment.

$$PL(dB) = 40\log(d) - 7.04 + \text{Clutter factor (dB)} \tag{5.3}$$

The ground reflection model with different levels of clutter has been compared with the Okumara-Hata urban model [108]. Figure 5.2 shows that the ground reflection model (GRM) with 30dB clutter gives a similar path loss to the Okumara-Hata urban model. In a typical cellular environment, the heights of the base station and the mobile station are about 15m and 1.5m respectively, however in this model, the transceivers are assumed to be hand held, so the height of the transmitter and receiver antennas is 1.5m. This reduces the gain by:

$$10\log(15)^2 - 10\log(1.5)^2 = 20\text{dB} \tag{5.4}$$

Hence, the propagation model for two hand held transceivers is assumed to be:

$$PL(dB) = 40\log(d) - 7.04 + 50dB \tag{5.5}$$

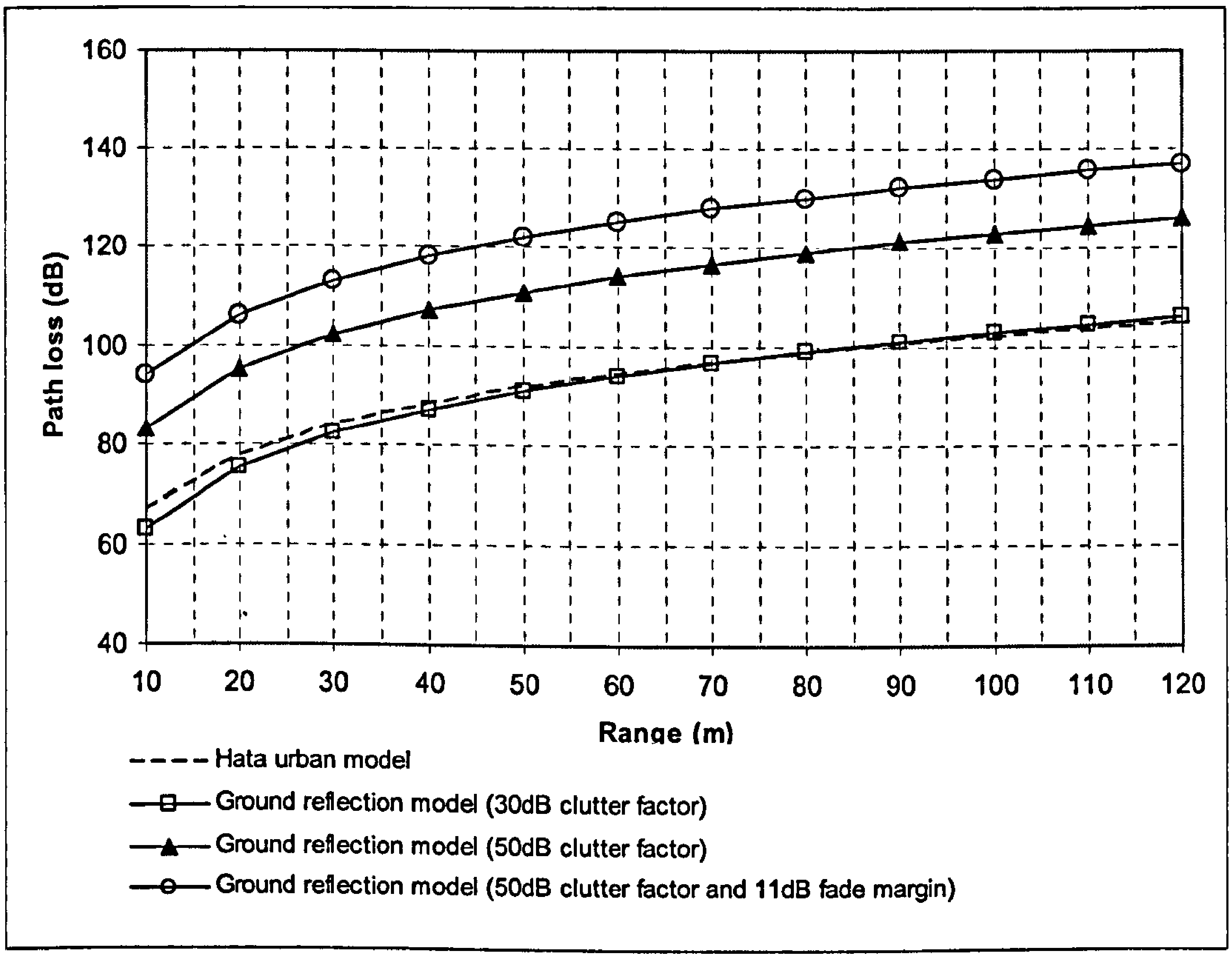


Figure 5.2 Comparison of path loss models

If a network is deployed in a built environment that consists of buildings, then communication takes place via scattering of electromagnetic waves from surfaces or diffraction over and around buildings. This is due to the fact that the height of the transmitter and receiver antennas are well below the surrounding buildings. As a result, there is multi path propagation and the signal undergoes slow and fast fading. In order to provide a reliable link, an extra fade margin is used in the path loss equation. If a user is moving at walking speed (3 km/hr), then 7dB for slow fade and 4dB for fast fade margin is assumed for an outdoor environment [108]. This assumes that fading is also compensated by controlling transmission power with the help of beacon frames.

$$PL(dB) = 40\log(d) - 7.04 + 50dB + 11dB \tag{5.6}$$

Figure 5.2 shows the relationship between range and path loss for the propagation model used in this simulation.

5.4.1.3 Beacon channel power budget

The data channel has a fixed maximum transmission power of +20dBm according to the IEEE 802.11 specification [12]. The maximum path loss for this transmission is 100dB using the receiver sensitivity of -86dBm. The calculations are given in Appendix B.

The power budget for the beacon channel is based on a path loss of 106dB. The clutter factor of 50dB is also included in the path loss which gives a maximum transmission range of 20m. For a beacon rate of 32kbps, this corresponds to a beacon transmission power of 500uW (-3dBm), see Appendix C. As only one terminal is transmitting at one time, the average transmission power per node is 500μW/N, where N is the number of nodes. So for a 25 node network, the transmission power of the beacon frame per node is about 20μW (-47dBm). By calculating the received power of the beacon frame, the minimum transmission power between two nodes can be calculated for the data transmission.

In contrast, the power consumption of beacon frame per node is about 20μW at the same path loss. Therefore, beacon power is small in comparison to the power consumption of the data channel and will not have a significant effect on the overall power consumption of each node.

5.4.1.4 Beacon channel bit rate

The rate at which the control data is updated depends on the speed of the mobile terminals. The faster the nodes are moving, the more frequently the control data must be transmitted. In addition, the reachability of the beacon frames also depends upon the topology of the network. For a network size of 50m x 50m, the maximum distance between the nodes is about 70m, see figure 5.3. However, based on the 106dB path loss and 50dB clutter factor, the maximum transmission range of the beacon frame is 20m. Hence three hops are needed to ensure the information is distributed using the beacon frames to all the nodes.

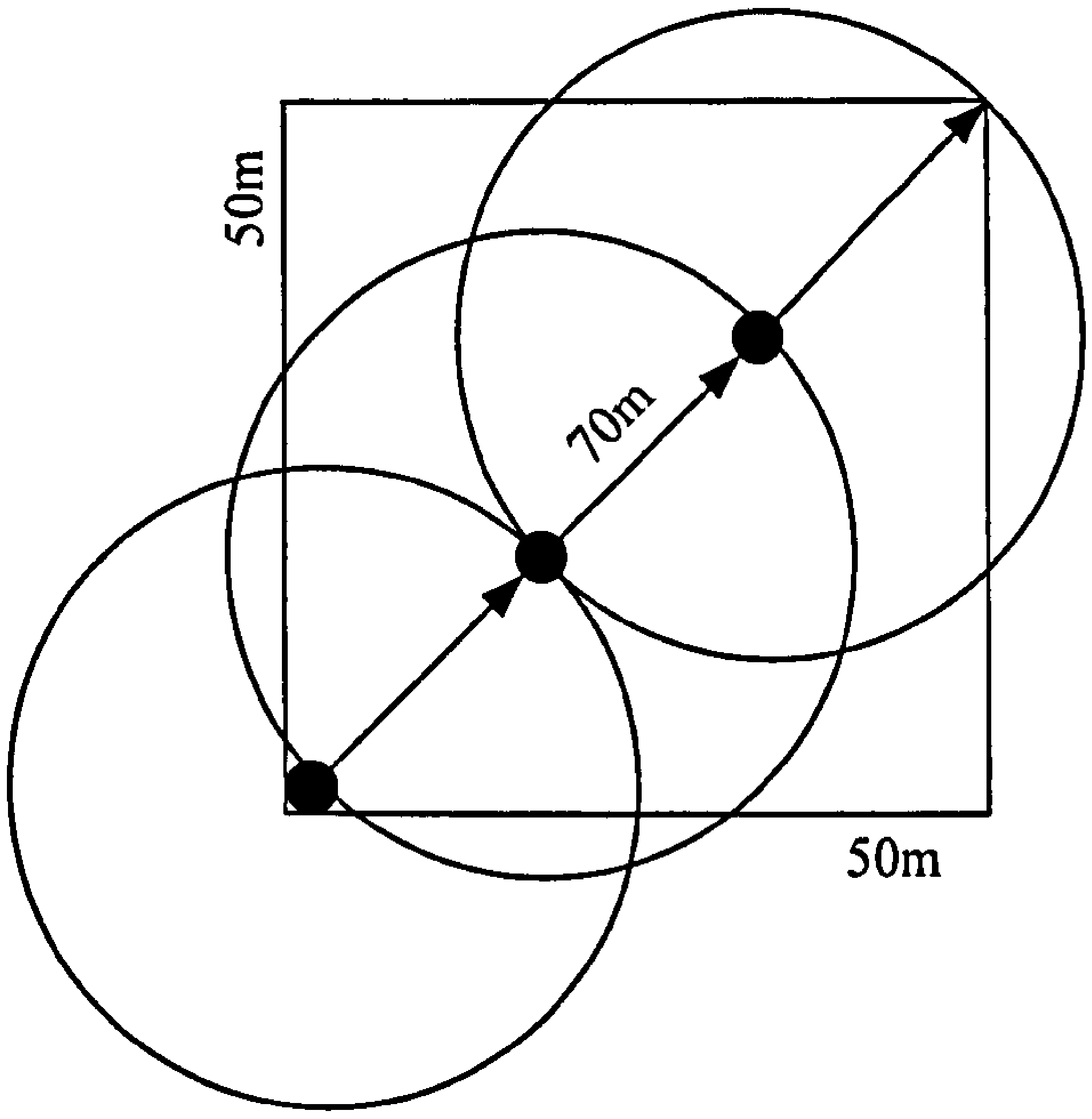


Figure 5.3 Three broadcasts required for the network size of 50m x 50m

If the speed of the node is V km/hr and the maximum distance that a node can move without a significant change to the topology is D meters. The minimum bit rate (bps) of the beacon channel is given by:

$$\text{Beacon bit rate} = 6.66\,N\,(4+2N)\;V\,/\,D \tag{5.7}$$

The calculation of the beacon bit rate is described in Appendix D. Figure 5.4 shows a graph of the beacon bit rate as a function of the number of nodes in the network for node velocities of 3km/hr, 20km/hr and 100km/hr. The beacon information must be distributed before a node has moved 5m (an assumption based on the size of the network and the density of nodes). It can be seen that for a 25 node network, the required beacon bit rate for a velocity of 3km/hr is approximately 5kbps. This rises to 36kbps for a node velocity of 20km/hr and 180kbps for a node velocity of 100km/hr. The beacon rate must be pre-selected for the network based on the maximum expected speed of the nodes using the network. This rate should be selected such that the movement of the nodes within the beacon period is not sufficient to affect the route selection.

There are a total of $N(4+2N)$ bytes in the full frame which carries all the node information, corresponding to 10,800 bits in a 25 node network. Each node re-broadcasts the most recent information it has received. It may take up to 3 frames for information to be distributed to all nodes in the network depending on the attenuation between the nodes. So for a bit rate of 5kbps, it will take up to 6.48 seconds to establish a route.

A maximum value for the time to select a route should be set for the network. This may depend on the application, but assume it is 1 second, then the beacon bit rate will need to be increased to 32kbps.

For this delay and for node speeds up to about 20km/hr, the bit rate is limited by the time to establish a route. For higher speeds, the beacon bit rate may need to increased depending on the number of nodes in the network to maintain the accuracy of the topology and the correct attenuation between the nodes. Figure 5.4 shows that for a speed of 100km/hr, a bit rate of 32kbps is sufficient only for up to 10 nodes in the network.

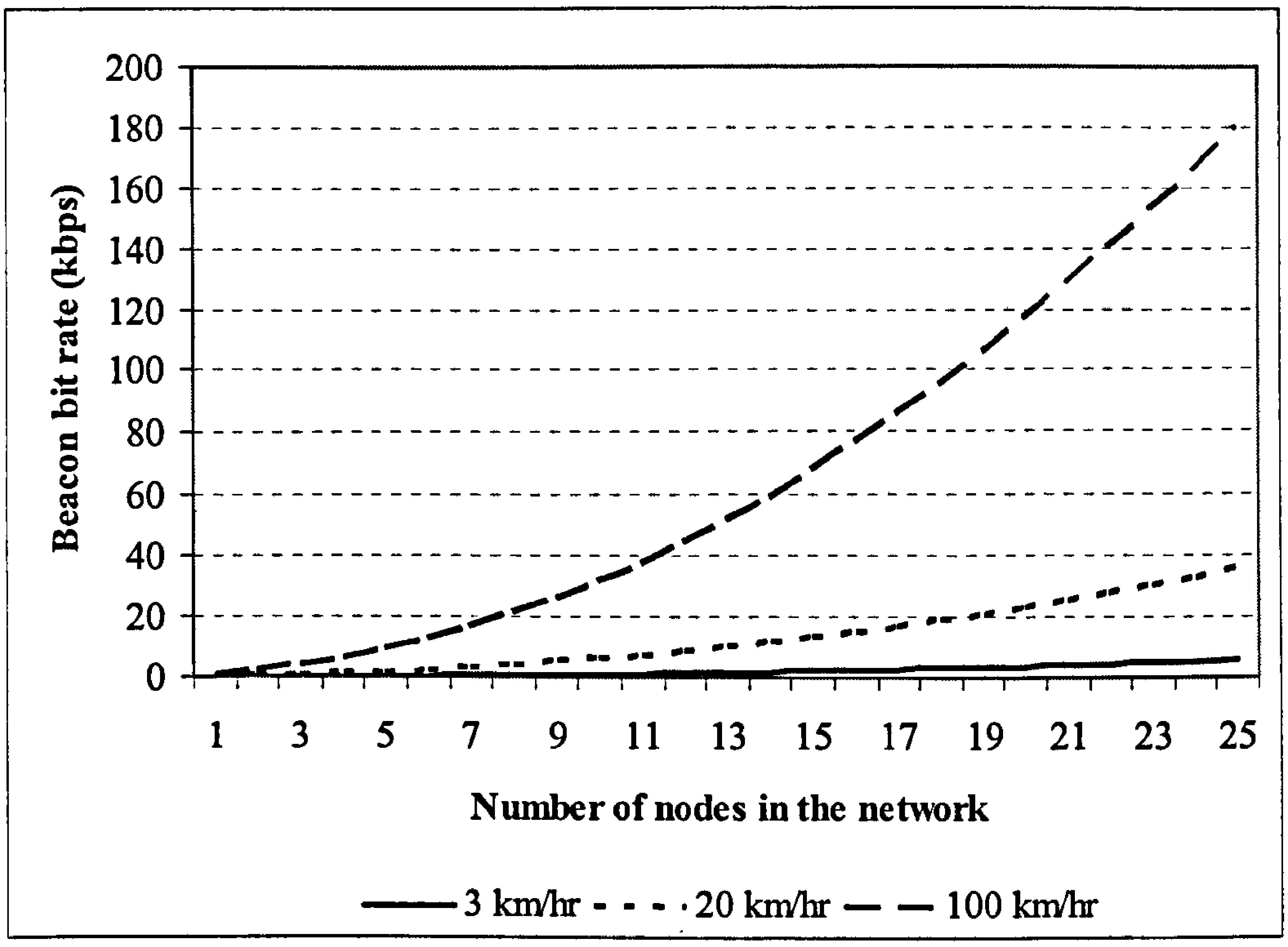


Figure 5.4 Beacon bit rate as a function of the number of nodes in the network

The criteria for a node to select an on-demand resource discovery is based on the mean time between transmissions from a node. The mean time between transmissions should be greater than the time to establish a route for on-demand operation to provide a saving in node battery consumption. If the time taken to select a route is 1 second for a beacon rate of 32kbps, then on-demand routing should only be selected if the average time between transmissions is greater than 1 second.

5.4.2 RTS/CTS

RTS/CTS frame uses a data channel and it is transmitted before the data transmission at the same power levels as data packets. The RTS frame is sent to reserve an end to end route. The RTS frame also carries the intermediate node IDs for the source-destination route. On receipt of an RTS, each node saves the incoming node ID and initiates RTS for the second relay node. When the destination node receives an RTS, it sends a CTS which propagates back through the same hops to the source node (the address field is reversed in the CTS frame). Figure 5.5 (a and b) shows the frame format of the RTS/CTS frame and figure 5.5 (c) shows the timing diagram. The process of RTS/CTS is described in Appendix E.

8	8		8	8	32
SA	RA-1	…….	RA-N	DA	FCS

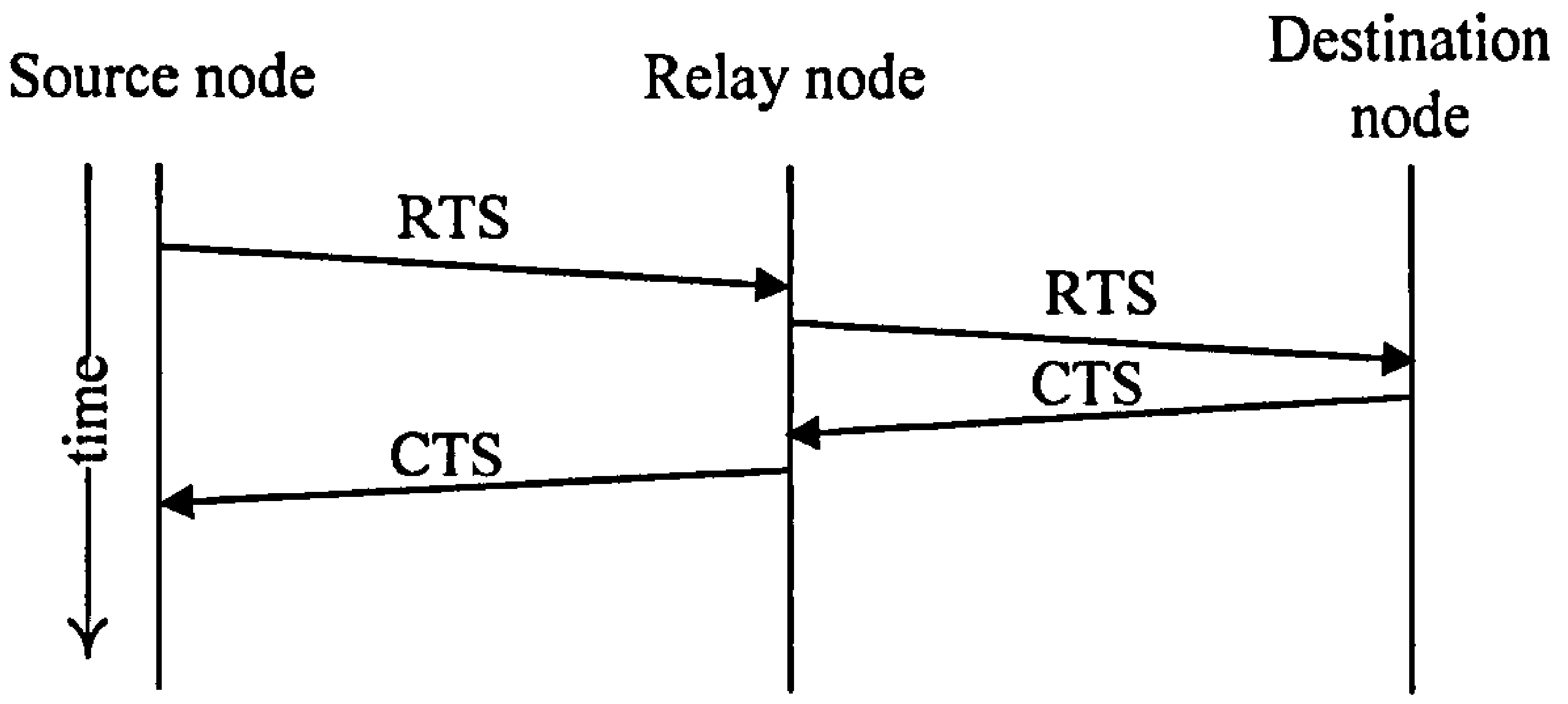
SA=source address, DA=destination address

RA-1=first relay address, RA-N=last relay address, FCS=frame check sequence

A) RTS frame format

8	8		8	8	32
DA	RA-N	…….	RA-1	SA	FCS

B) CTS frame format



C) RTS/CTS timing diagram

Figure 5.5 RTS/CTS frame format

5.5 Summary

This chapter describes the design of an ad hoc network with the objectives of maximising the network lifetime and minimising the power consumption. The design considers the physical layer based on the IEEE 802.11 specifications and uses single frequency for an end to end path; the medium access is based on a busy tone scheme and the source routing is used at the network layer.

In order to maximise the network lifetime and minimise the power consumption, the routing protocol depends on the information from all of the other nodes in the network. Such information is propagated with the help of beacon frames which are broadcasted in the network using an out-of-band signalling channel.

A two-ray ground reflection path loss model is used. The maximum path loss for the data channel is 106dB. Using the same path loss for the beacon channel, and the clutter factor of 50dB, the maximum transmission range of beacon frames are up to 20m.

For a beacon bit rate of 32kbps, the power consumption for beacon frame transmission per node is 20μW. The maximum data power per node is 100mW. The power consumption due to beacon power is small as compared the data transmission and will not have an effect on the overall power consumption of each node.

The beacon frame rate is dependent upon the velocity of nodes, the number of nodes, the minimum update distance, and the number of beacon frame transmission needed to ensure end to end connectivity.

RTS/CTS control frames are sent to reserve the end to end route, and to notify relay nodes about the selected route so that each node knows its next hop.

6. Simulation Model and Results

This chapter discusses the simulation model and presents the results of various routing schemes described in chapter 4. Section 6.1 discusses the simulation environment. Section 6.2 discusses how minimum transmission power is calculated. Section 6.3 discusses how the results are gathered. Section 6.4 presents results for a network using a point to point link. Section 6.5 presents results for a network using the minimum power routing scheme. Section 6.6 presents results for a network including a battery charge threshold. Section 6.7 presents results for the residual battery charge scheme. Section 6.8 presents results for the MPR/MBL approach. Section 6.9 discusses the impact on the network lifetime in different network scenarios. Section 6.10 compares the power consumption of the various routing schemes.

6.1 Simulation environment

In this model, it is assumed that each end to end route uses a single frequency. In addition, each node is either transmitting or receiving at one time, so data is transferred on a hop by hop basis. For example, in figure 6.1, if the source node A communicates with the destination node C via relay node B; then node B first buffers data from node A and when all the data has reached node B, then it is transmitted to node C.

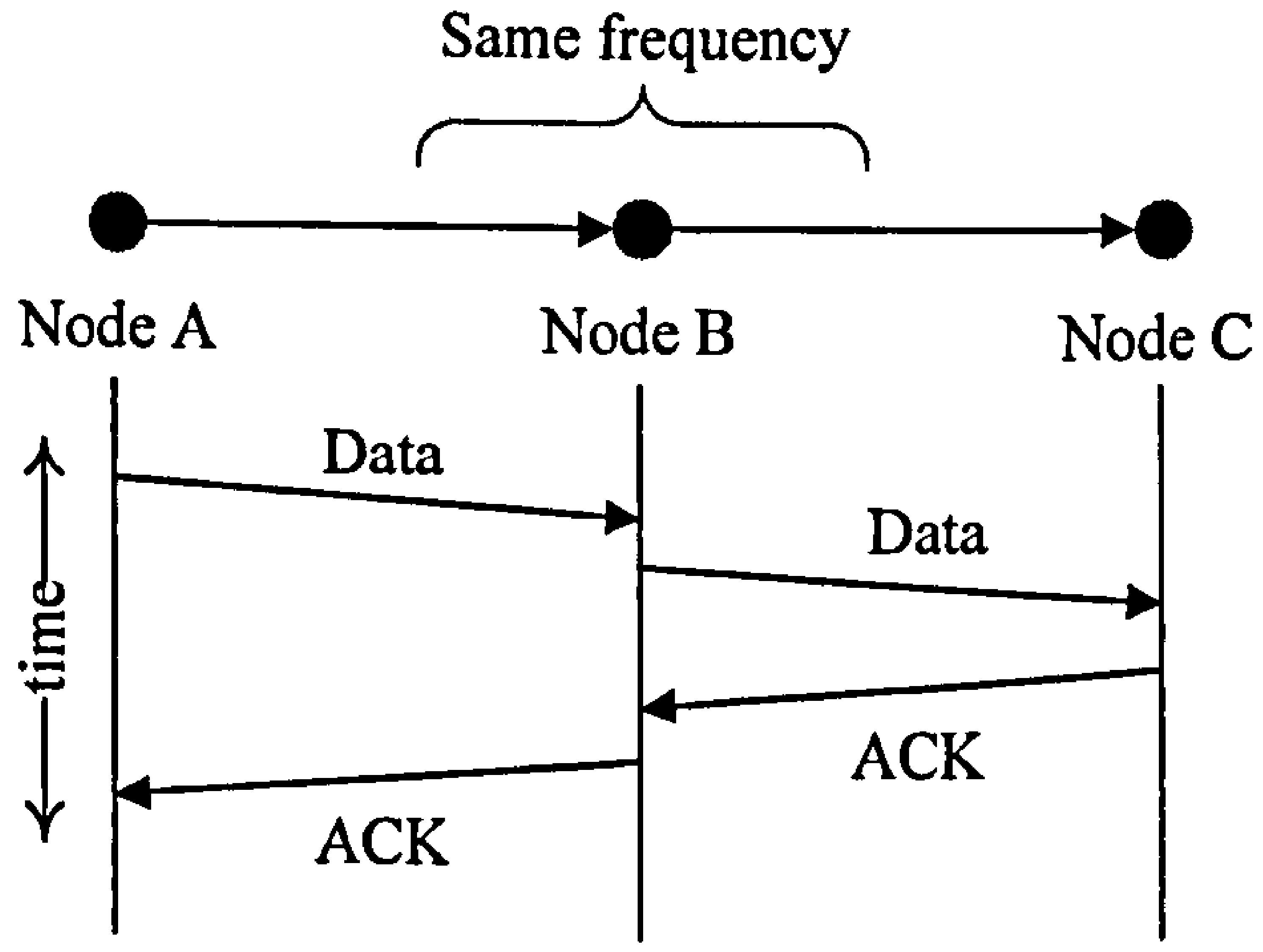


Figure 6.1 Communication between the nodes using a single frequency

Nodes are modelled to transmit 1 MB of data which could be voice/data/video. For example, in a rescue attempt, video footage is important to assess the extent of any damage. It is assumed that data is transmitted at 6 Mbps using Binary Phase Shift

Keying modulation (BPSK) which follows the IEEE 802.11 standard. An average receiver sensitivity value of -86 dBm is used based on commercially available datasheets [104, 109] for equipment supporting the IEEE 802.11 standard.

In order to calculate the network lifetime, the current drain due to transmitting, receiving and processing data for each transmission is calculated. The power required for processing is based on information from commercially available wireless transmission integrated circuit datasheets, and is in the range 400-723mW [103, 110, 111]. The transmission power is dependent upon the data rate, and the amplifier efficiency. The path loss is also included in the simulation model. It is assumed that all nodes are powered by Lithium-ion batteries of 700mAh charge which are commonly used in hand-held phones. At the start of the simulation, all nodes are fully charged and it is assumed that they cannot be recharged during the network operation, which reflects the environment in which they are expected to be used. In the model, it is assumed that all links are bidirectional in nature and the selected route from source to destination node will use same relay nodes in both directions. The model selects random source and destination nodes. Table 6.1 summarises the parameters which are used in the simulation.

Parameters	**Value**
Number of nodes	25
Network area	50m x 50m
Battery charge	700mA-hrs (fully charge)
Maximum allowed radiated transmission power	100mW [100, 101]
Receive power	710mW [103]
Processing power due to transmission	723mW [103]
Data rate	6 Mbps [100]
Receiver sensitivity	-86 dBm [104]
Modulation	BPSK [112]
Data size	1 MB
Antennas	Omni directional with unit gain
Mobility	Static
Power amplifier efficiency	40% [102]

Table 6.1 Simulation parameters of the modelled network

Figure 6.2 shows that 25 nodes are placed randomly in an area of 50m x 50m. A larger network area 100m x 100m is also analyzed, however due to insufficient power budget and number of nodes, larger network area experiences significant route blockages when modulation is adapted. The random attenuation between the nodes is shown in figure 6.3.

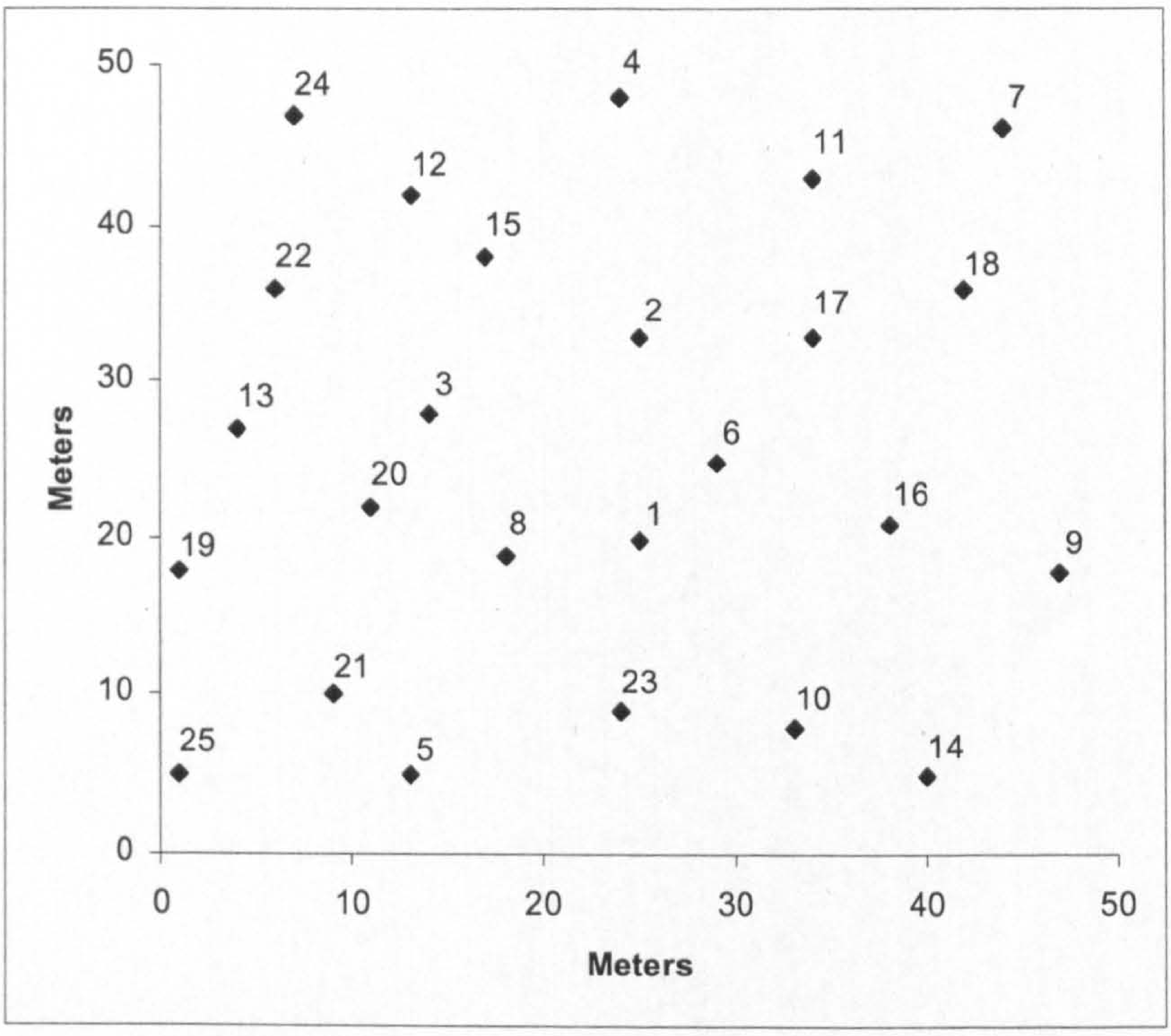


Figure 6.2 Node placement in the modelled network

Attenuation in dB

Nodes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1		48.5	47.3	47.6	44.6	44.8	48.9	31.5	48.8	49.1	48.5	36.6	46.2	49.4	49.0	45.7	49.8	49.4	37.5	49.8	45.6	47.2	48.8	37.3	47.7
2	48.5		46.7	44.7	47.9	48.1	44.2	44.5	49.2	49.2	47.7	49.9	49.6	43.6	48.4	49.9	43.9	47.3	40.3	50.0	48.3	32.0	47.6	40.0	40.1
3	47.3	46.7		49.0	44.5	36.6	44.7	45.8	44.8	49.8	49.9	46.0	44.4	42.1	42.1	48.1	46.1	46.2	48.5	45.1	48.0	43.2	42.7	47.7	39.1
4	47.6	44.7	49.0		46.6	49.6	44.2	48.9	45.8	44.6	49.6	48.0	48.0	46.3	39.9	47.5	48.4	49.6	49.2	33.5	47.4	49.6	46.3	48.3	47.0
5	44.6	47.9	44.5	46.6		47.1	46.7	45.5	46.1	44.3	37.5	43.9	49.9	37.8	45.9	45.6	46.9	41.9	46.8	44.1	48.0	47.3	41.9	49.7	48.2
6	44.8	48.1	36.6	49.6	47.1		47.0	45.9	40.3	48.9	46.6	48.8	47.8	49.2	32.7	43.2	38.7	40.2	45.2	41.1	14.0	47.3	48.2	47.4	49.2
7	48.9	44.2	44.7	44.2	46.7	47.0		39.1	42.8	48.3	46.6	45.5	41.8	48.5	49.7	47.2	39.5	48.8	46.0	46.6	46.9	43.2	45.2	39.8	47.7
8	31.5	44.5	45.8	48.9	45.5	45.9	39.1		42.3	49.7	39.9	46.5	44.4	49.4	48.8	44.4	48.3	44.1	39.5	34.9	45.1	49.0	44.7	43.7	46.8
9	48.8	49.2	44.8	45.8	46.1	40.3	42.8	42.3		44.1	45.3	36.5	46.8	43.1	49.4	47.7	48.8	49.7	45.2	47.3	39.1	48.0	46.1	49.8	40.6
10	49.1	49.2	49.8	44.6	44.3	48.9	48.3	49.7	44.1		49.7	47.9	45.4	41.7	46.8	43.4	50.0	41.2	34.6	45.4	47.4	49.7	47.3	46.1	49.3
11	48.5	47.7	49.9	49.6	37.5	46.6	46.6	39.9	45.3	49.7		49.2	48.3	48.6	50.0	45.3	46.9	46.2	48.4	42.5	46.3	47.3	49.1	47.3	46.3
12	36.6	49.9	46.0	48.0	43.9	48.8	45.5	46.5	36.5	47.9	49.2		47.1	43.6	47.9	46.9	48.3	49.5	45.7	44.8	44.7	41.8	47.2	43.5	47.7
13	46.2	49.6	44.4	48.0	49.9	47.8	41.8	44.4	46.8	45.4	48.3	47.1		45.6	49.4	46.8	42.8	48.4	48.7	47.9	48.9	42.1	49.1	43.1	49.8
14	49.4	43.6	42.1	46.3	37.8	49.2	48.5	49.4	43.1	41.7	48.6	43.6	45.6		38.2	37.9	49.0	45.8	46.7	40.8	40.6	42.4	36.8	48.5	47.3
15	49.0	48.4	42.1	39.9	45.9	32.7	49.7	48.8	49.4	46.8	50.0	47.9	49.4	38.2		47.5	43.4	46.7	48.7	48.8	46.0	49.6	48.7	39.5	48.0
16	45.7	49.9	48.1	47.5	45.6	43.2	47.2	44.4	47.7	43.4	45.3	46.9	46.8	37.9	47.5		48.5	32.0	46.3	46.0	44.4	49.9	49.0	48.4	46.2
17	49.8	43.9	46.1	48.4	46.9	38.7	39.5	48.3	48.8	50.0	46.9	48.3	42.8	49.0	43.4	48.5		48.7	44.4	45.5	46.4	49.8	40.8	48.1	45.4
18	49.4	47.3	46.2	49.6	41.9	40.2	48.8	44.1	49.7	41.2	46.2	49.5	48.4	45.8	46.7	32.0	48.7		40.2	42.7	38.9	46.4	49.8	47.3	46.9
19	37.5	40.3	48.5	49.2	46.8	45.2	46.0	39.5	45.2	34.6	48.4	45.7	48.7	46.7	48.7	46.3	44.4	40.2		49.9	43.4	45.8	46.0	44.5	47.0
20	49.8	50.0	45.1	33.5	44.1	41.1	46.6	34.9	47.3	45.4	42.5	44.8	47.9	40.8	48.8	46.0	45.5	42.7	49.9		41.4	47.1	49.8	47.5	49.6
21	45.6	48.3	48.0	47.4	48.0	14.0	46.9	45.1	39.1	47.4	46.3	44.7	48.9	40.6	46.0	44.4	46.4	38.9	43.4	41.4		48.2	46.4	48.4	38.1
22	47.2	32.0	43.2	49.6	47.3	47.3	43.2	49.0	48.0	49.7	47.3	41.8	42.1	42.4	49.6	49.9	49.8	46.4	45.8	47.1	48.2		48.8	48.5	47.0
23	48.8	47.6	42.7	46.3	41.9	48.2	45.2	44.7	46.1	47.3	49.1	47.2	49.1	36.8	48.7	49.0	40.8	49.8	46.0	49.8	46.4	48.8		41.9	43.5
24	37.3	40.0	47.7	48.3	49.7	47.4	39.8	43.7	49.8	46.1	47.3	43.5	43.1	48.5	39.5	48.4	48.1	47.3	44.5	47.5	48.4	48.5	41.9		45.1
25	47.7	40.1	39.1	47.0	48.2	49.2	47.7	46.8	40.6	49.3	46.3	47.7	49.8	47.3	48.0	46.2	45.4	46.9	47.0	49.6	38.1	47.0	43.5	45.1	

Figure 6.3 Random attenuation in dB between the nodes

6.2 Minimum transmission power

Minimum transmission power between two nodes is calculated using the following equation:

Minimum Transmission Power (dBm) = Path loss (dB) + Receiver sensitivity (dBm) (6.1)

where

path loss is calculated with the help of equation 5.5 (chapter 5) and the receiver (Rx) sensitivity is taken from the vendor's datasheet [109] using receivers based on the IEEE 802.11 specifications [12].

6.2.1 Route calculation

The route is calculated according to the routing scheme. There are four routing schemes used in the simulation model; a) minimum power routing, b) minimum power routing with a battery charge threshold scheme, c) residual battery charge, and d) MPR/MBL scheme. The steps of the simulation process are shown in Appendix F, G and H. The Bellman-Ford algorithm is used in these routing schemes. The working of this algorithm using a 7-node example model is illustrated in Appendix I.

6.3 Simulation procedure

The network has been modelled using the topology shown in figure 6.2. The network lifetime is the time over which all nodes in the network have sufficient battery charge to actively participate in communication with other nodes. The network lifetime is analysed for a network with the simulation parameters defined in table 6.1. Source and destination nodes are selected randomly. All the nodes initially have full battery charge (700mAhrs). The network lifetime is calculated by analysing the current taken from the battery in each node. The source node transmits data size of 1MB at 6Mbps. This corresponds to the data transmission time of 1.4 seconds. The transmitter current drain depends upon the transmit power (the power is calculated with the help of path loss between the nodes) and the power amplifier efficiency of the transmitter. The Bellman-Ford Algorithm is executed whenever there is a change in the network topology or change in the parameters which are used during route calculations. For example, in the minimum power routing, the change in the network topology occurs when a node reaches zero battery charge. In the battery charge threshold scheme, the topology change occurs when the nodes reaches threshold and are isolated as relay nodes. In the residual battery charge and MPR/MBL schemes, the topology does not change however one of the parameters for route selection is 'residual battery charge which changes with time and routing algorithm is executed before the data transmission in order to have accurate routes to the destination nodes.

6.4 Results (point to point links)

In this approach, the source node communicates directly to the destination node. Due to the allowed transmitter power of +20dBm, some of the destination nodes are not reachable, and therefore some of routes are blocked. Figure 6.4 shows the battery charge of each node in the network as a function of time. The first node has reached zero battery charge is 26 hours; however 39% of the routes are blocked.

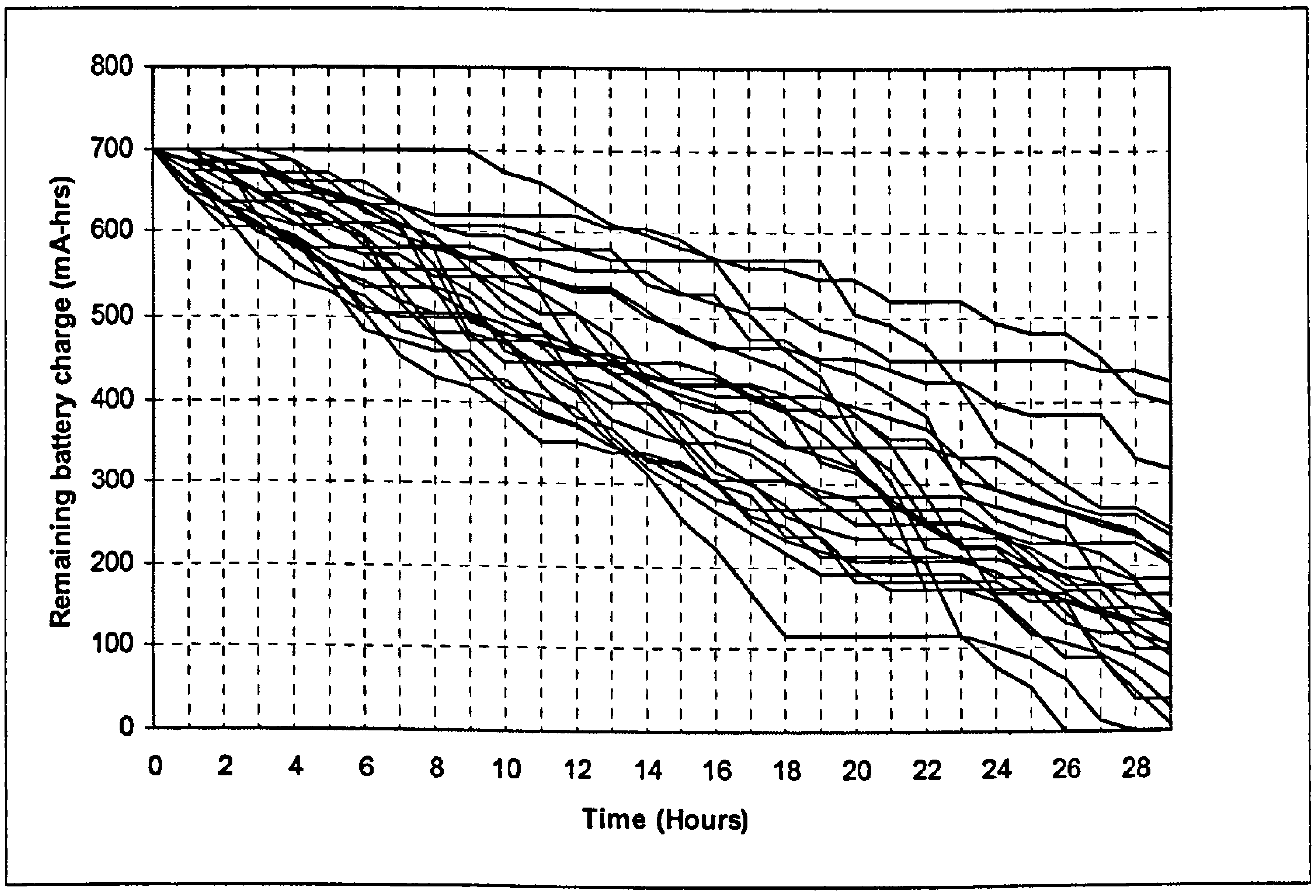


Figure 6.4 Battery charge vs time – point to point links (with power limitation)

The network lifetime is calculated for the point to point link with no transmitter power restriction in order to overcome route blockages. In the simulation model, all of the routes are directly connected with no route blockages. The maximum transmission power is reached up to +38dBm. Figure 6.5 shows that the use of high transmitter power reduces the network lifetime significantly.

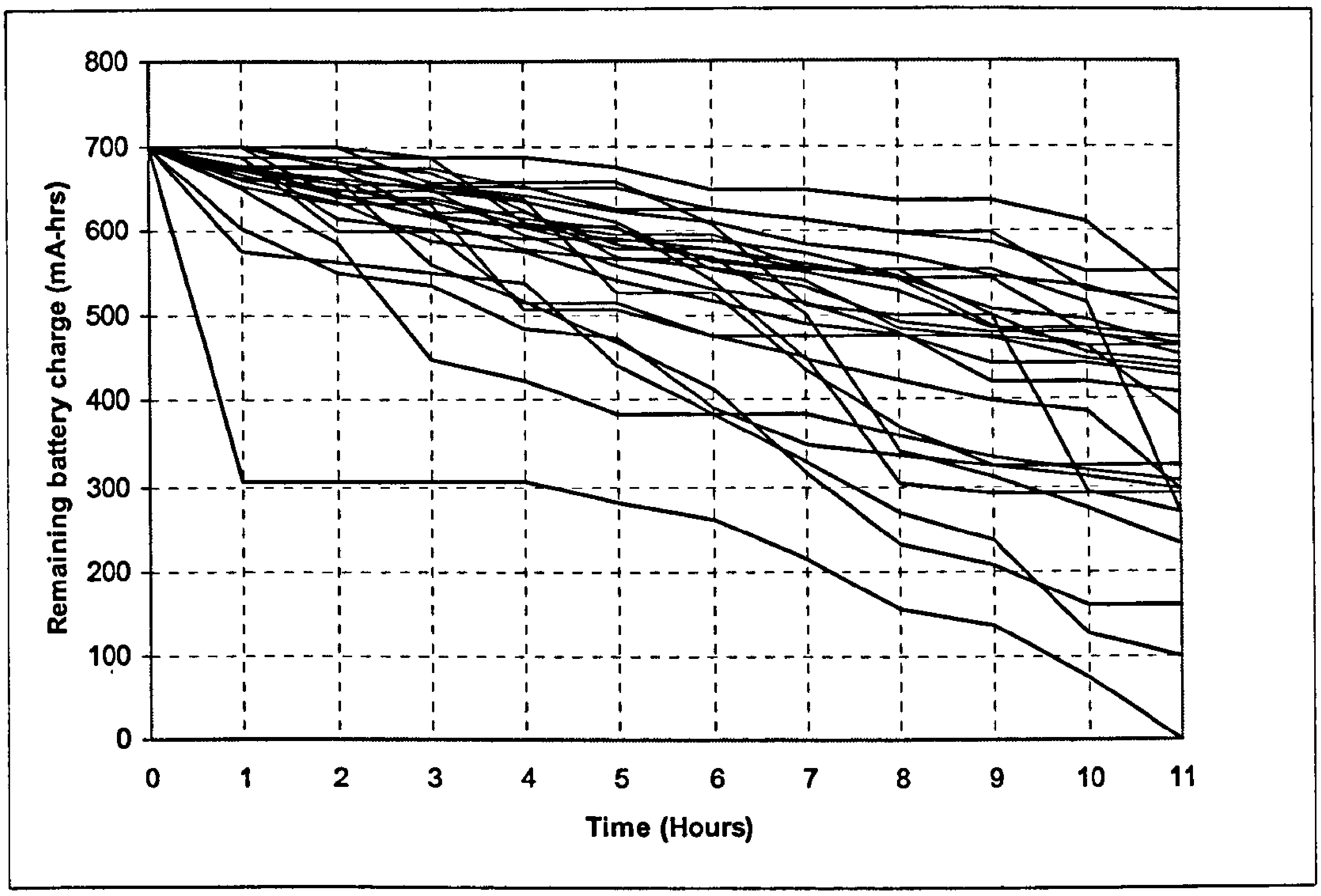


Figure 6.5 Battery charge vs time – point to point links (without power limitation)

6.5 Results (minimum power routing)

In this approach, all 25 nodes are available all of the time to relay data in addition to their normal function as a source and destination node. As discussed in chapter 4, section 4.4.1, some nodes may be used more than others due to their location in the network. Figure 6.6 shows the battery charge of each node in the network as a function of time, assuming all batteries initially have a full charge. It can be seen that node 6 is the most heavily used node because it is selected as a relay for many of the routes and therefore its battery is the first to expire. Node 22 is the least used node. This is because this node is located at the edge of the network and is therefore less likely to be used as a relay node, see figure 6.2. The battery usage of other nodes is spread between these extremes. The variation in the probability of usage of each node leads to a spread in the times taken for the battery charge of each node to reach zero. The lifetime is limited by the most heavily used node with the first node having zero battery charge is 9 hours, while the batteries of other nodes have battery cnarge remaining.

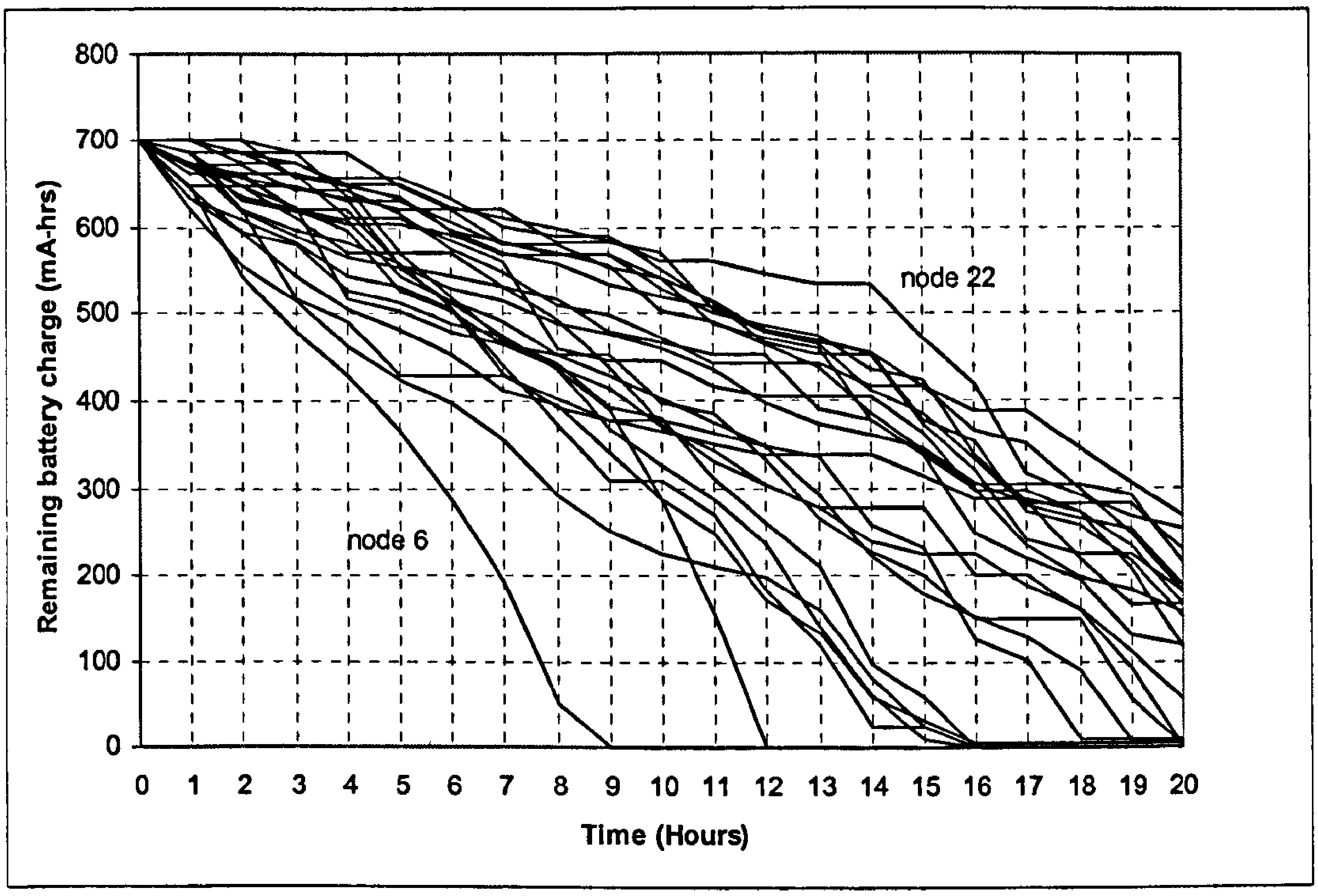


Figure 6.6 Battery charge vs time – minimum power routing

6.6 Results (battery charge threshold approach)

6.6.1 Lower threshold – 10% battery charge threshold approach

The concept of using a battery threshold was introduced in the section 4.4.2, chapter 4. The aim is to balance the usage of heavily used nodes by withdrawing them from use as a relay when their battery charge reaches a pre-defined level. Using the network topology shown in figure 6.2, figure 6.7 shows the remaining battery charge against time with a threshold set at 10% of the maximum battery charge. For clarity, the nodes with the lowest and highest battery consumption are shown and two intermediate nodes. It can be seen that the most heavily used node reaches threshold first and then it only functions as a source and destination node and not a relay node. The minimum power route is calculated using the available nodes.

As the number of nodes that have reached threshold increases, the current required for transmission by the remaining nodes increases. With a threshold setting of 10%, the time taken for the first node to have zero battery charge is about 14 hours which is a 56% improvement over operation with MPR. With the topology used for this simulation, node 8 is the first node to reach threshold.

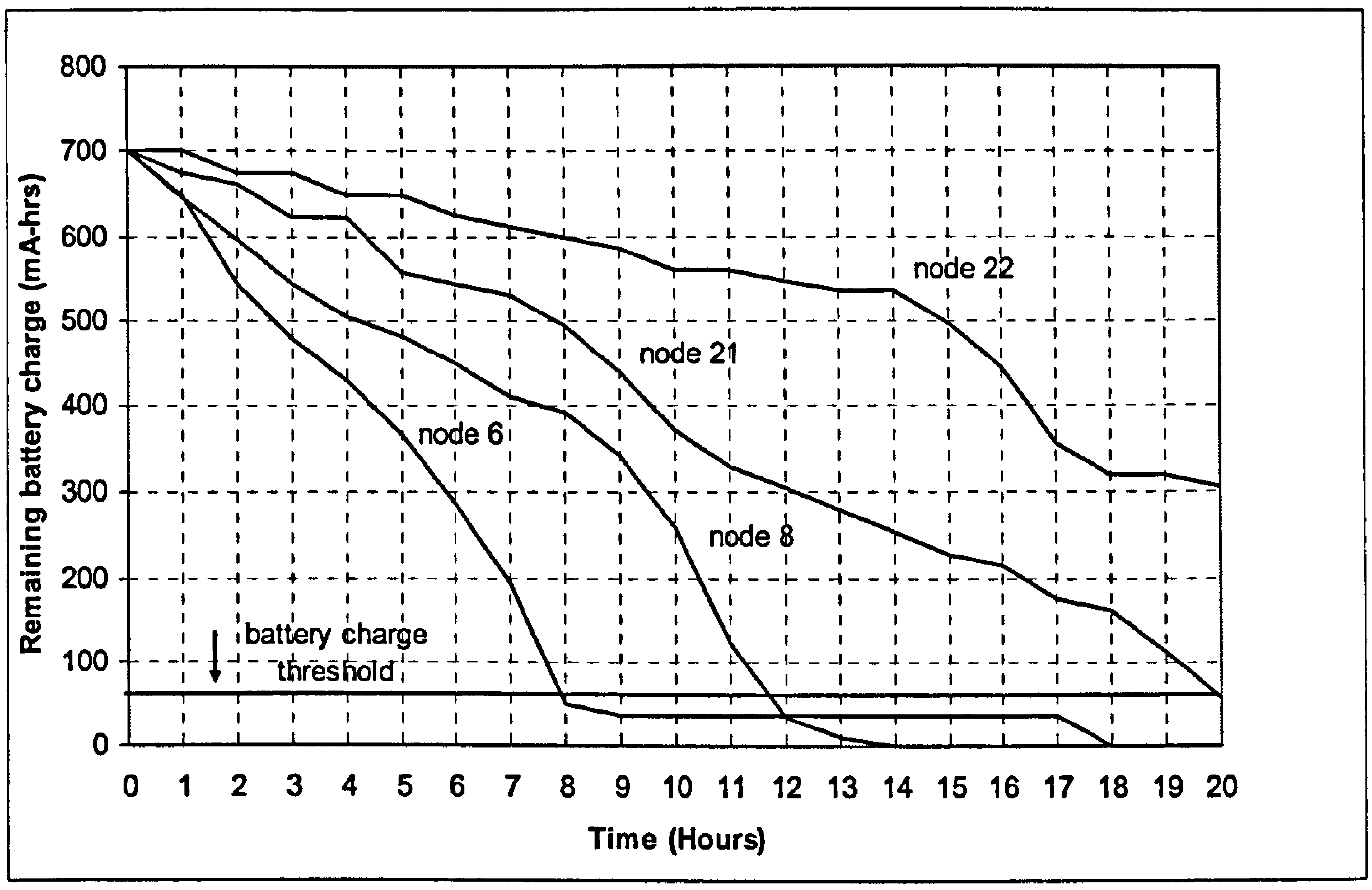


Figure 6.7 Battery charge vs time (threshold at 10% of maximum battery charge)

6.6.2 Higher threshold – 90% battery charge threshold approach

When higher battery charge threshold is used, then nodes are withdrawn from use as a relay at an early stage. As a result, a route between the source and destination routes may not exist because of the lack of relay nodes in the network, and hence route blockage exists. In order to prevent route blockage, isolated relay nodes need to be reintroduced.

Figure 6.8 shows the remaining battery charge against time with a threshold set at 90% of the maximum battery charge. In this case the battery of the most heavily used node (node 6) reaches 90% of its maximum battery charge after approximately one hour and is withdrawn from use as a relay. When fourteen nodes have reached threshold, which occurs after 2.5 hours, some routes become blocked. This means that it is not possible to establish a route from a source to a destination node with the remaining nodes with the maximum power limited to 100mW. This is shown as a vertical marker on the 90% threshold line in figure 6.8. As the key criteria is to enable all nodes to be able to communicate, then when a blockage occurs, all nodes are restored to operate as relay nodes. After this, the minimum power route is again selected from all the available nodes and node 6 again is most heavily used and reaches zero battery charge about 11 hours this represents a 22% improvement over the case with MPR but lower than the case with a 10% threshold.

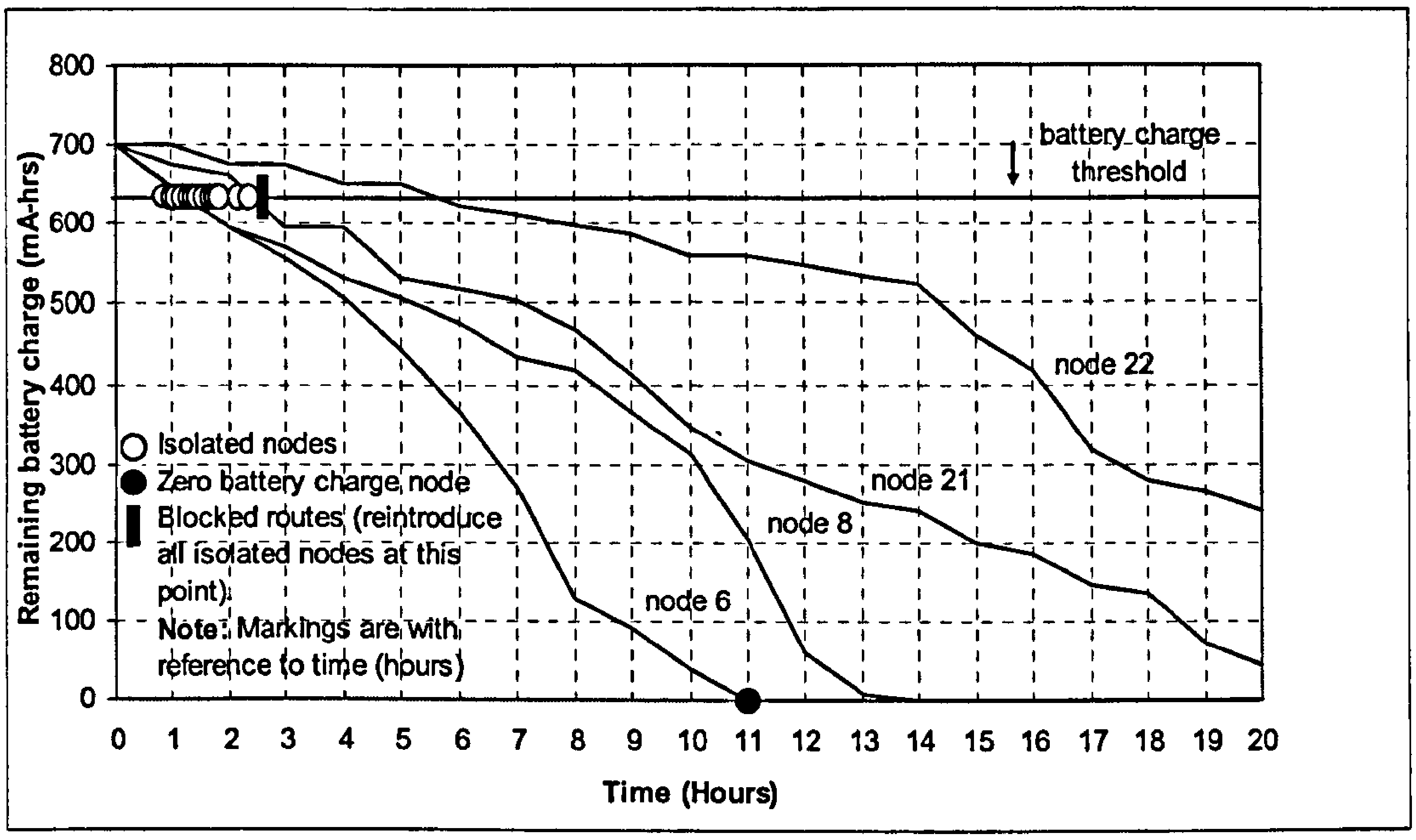


Figure 6.8 Battery charge vs time (threshold at 90% of maximum battery charge)

6.6.3 Optimum threshold

The investigations described in sections 6.6.1 and 6.6.2 shows that the network lifetime can be increased by using a battery charge threshold scheme. However the results indicate that the maximum network lifetime varies with different thresholds. Further tests have been carried out to investigate the optimum threshold. Ten different network topologies (50m x 50m) are analysed. In each network topology, twenty simulations have been carried out and in each simulation, battery charge threshold is increased by 5%. Figure 6.9 shows the network lifetime as a function of the battery charge threshold and it shows that the maximum network lifetime is achieved when the threshold is set at 30% of the maximum battery charge.

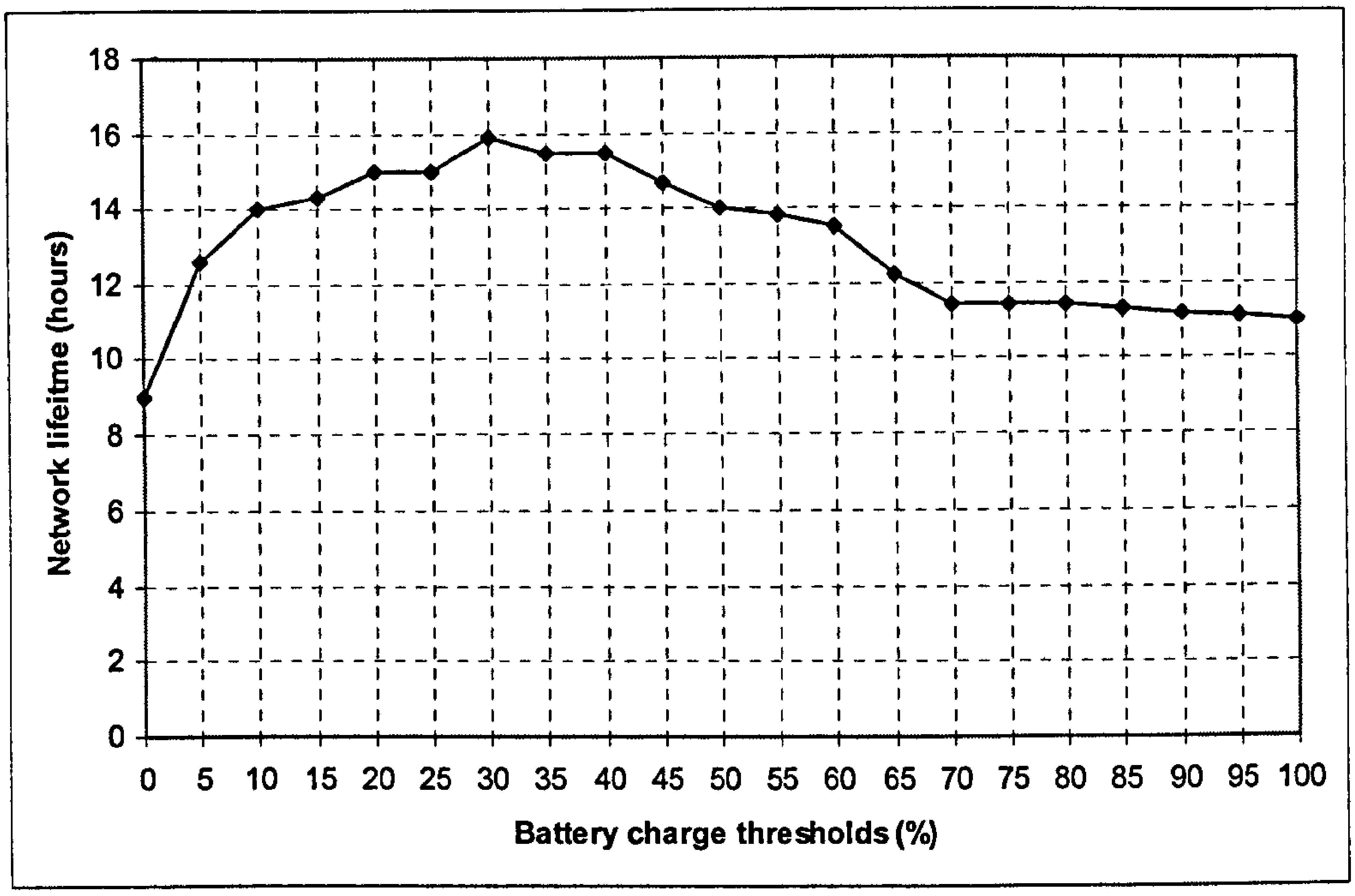


Figure 6.9 Network lifetime at various battery charge thresholds

Figure 6.10 shows the remaining battery against time with an average threshold set at 30% of the maximum battery charge. For this threshold, route blockage is experienced due to a lack of sufficient relays after 15 hours. At that time all nodes are returned to function as relays and it is node 8 that first reaches end of life. This occurs after 16 hours which represents a 78% improvement over the case with MPR.

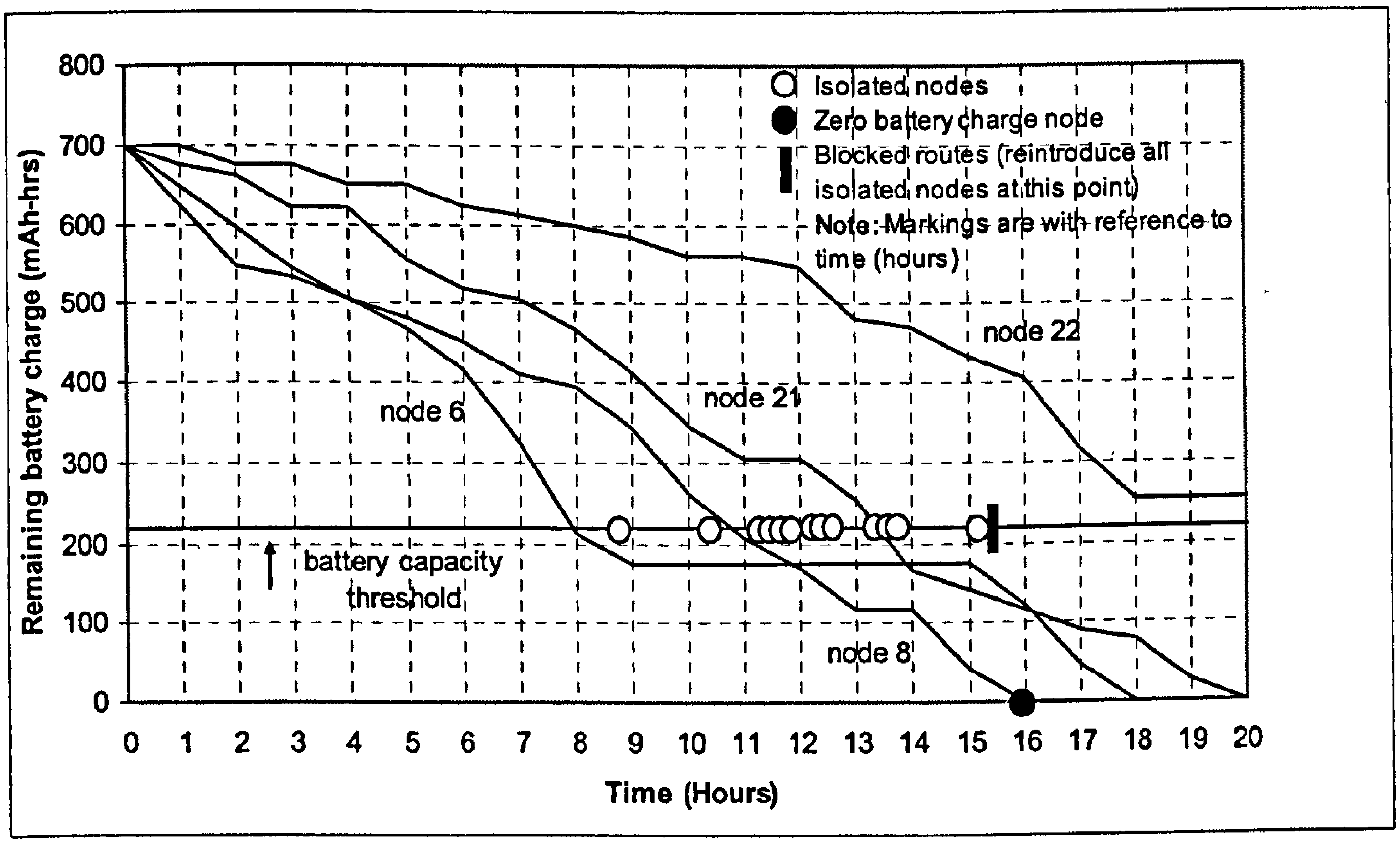


Figure 6.10 Battery charge vs time (optimum threshold at 30% of maximum battery charge)

6.7 Results (residual battery charge scheme)

Simulation experiments have also been carried out using a routing scheme based only on selecting routes with the maximum residual battery charge. The same network model, the same topology and simulation parameters as described in table 6.1 have been used. Figure 6.11 shows that the nodes (node 6 and node 8) that were heavily used in the minimum power and battery charge threshold schemes are used more evenly in the residual battery charge routing scheme. This scheme gives a network lifetime of about 20 hours.

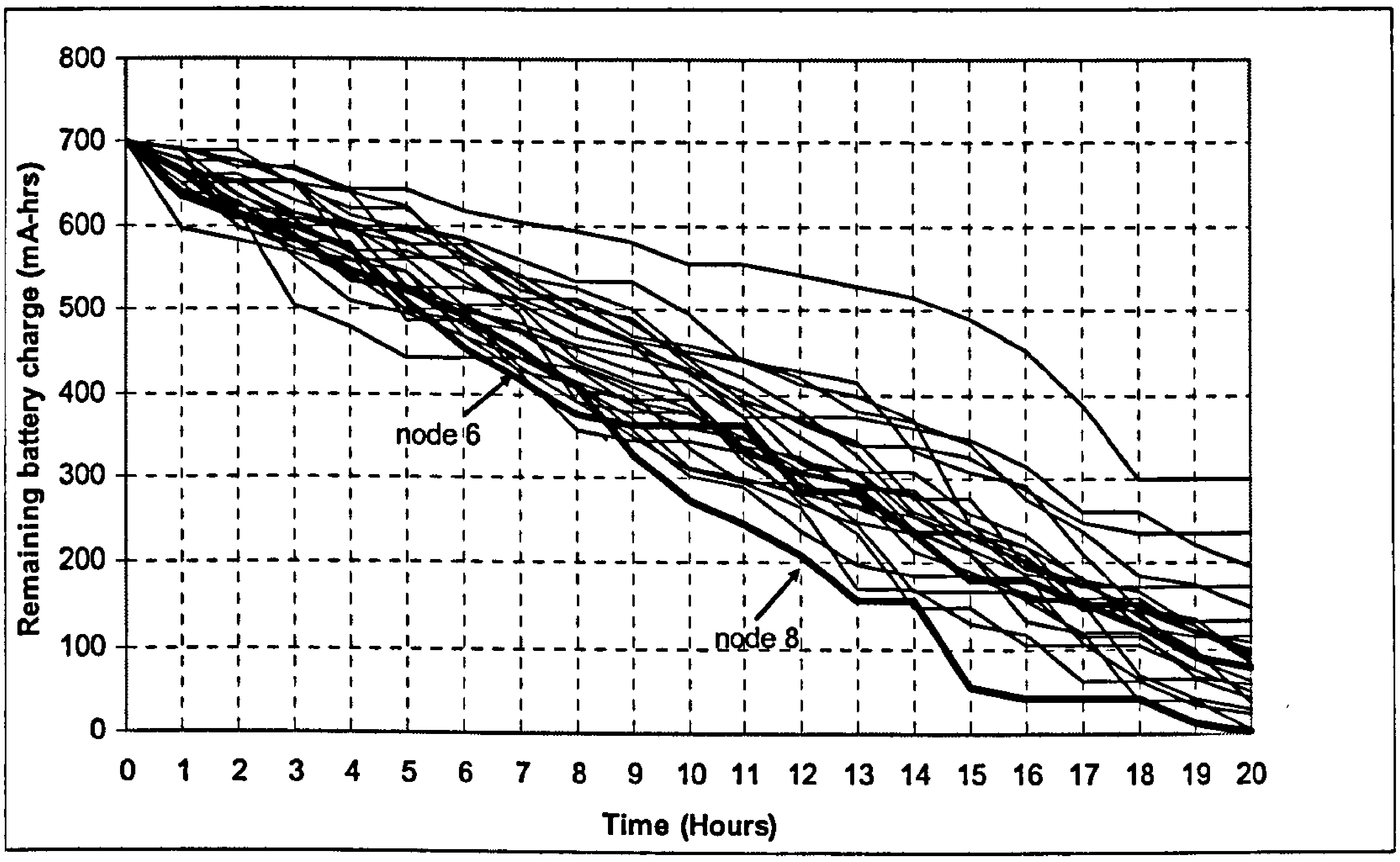


Figure 6.11 Battery charge vs time – residual battery charge routing

6.8 Results (MPR/MBL)

Simulations have been carried out to analyse the performance of the MPR/MBL approach described in section 4.6 using the simulation parameters described in table 6.1.

6.8.1 MPR/MBL with linear battery weighting

Simulations have been carried out for the case defined in section 4.6.1 where the cost associated with the battery charge is linearly related to the residual battery charge. The average transmission power for 25-node simulation model is calculated which is 23mW. This power is used to identify cross over points. Table 6.2 shows the network lifetime when cross over point is 200mAhrs and 600mAhrs respectively.

6.8.2 MPR/MBL with inverse battery weighting

Simulations were carried out for the case defined by equation 4.12 where the cost associated with the battery charge is inversely related to the residual battery charge.

6.8.2.1 MPR/MBL at cross over point (200mAhrs)

When a low cross over point (see equation 4.12, section 4.6.2) such as 200mAhrs is used, then a node which has a battery charge less than 200mAhrs is avoided as a relay node. However, as discussed in section 4.6.2 that the route A which consist of a low battery node is selected even when the available battery charge is less than the cross over point of 200mAhrs when a=1. Table 6.2 shows that the network lifetime is 10 hours when MPR/MBL with inverse battery weighting at cross over point 200mAhrs with a=1 is used. When exponent order is increased (for example a=5), then the battery term rapidly increases when residual battery charge reduces below the cross over point. Hence low battery nodes are avoided as relays, and therefore network lifetime is increased to 14 hours over the case with a=1.

6.8.2.2 MPR/MBL at cross over point (600mAhrs)

When a high cross over point (see equation 4.12, section 4.6.2) such as 600mAhrs is used, then at low exponent order (a=1), the network lifetime is 13 hours. The improvement in the network lifetime as compared with 200mAhrs cross over point (a=1) case is due to the fact that the nodes are avoided as relays at an early stage, and enough battery left to become a source or a destination node.

At low exponent order (a=1), low residual battery charge nodes are not avoided as relays. This case is already discussed in section 4.6.2. The simulation result in figure 6.12 shows that nodes are not evenly utilised and the network lifetime is limited by the overused node (node 8). When exponent order is increased (for example, a=5), then the battery term in equation 4.15 increases rapidly when residual battery charge reduces below the cross over point of 600mAhrs (see figure 4.13). Figure 6.13 shows that node 8 is prevented from use as a relay node. In addition, it can be seen that all nodes are more evenly used as compared to a=1.

MPR/MBL scheme (linear and exponential relationships)	Lifetime (Hrs.)
Linear, 200mAhrs cross-over point	10
Linear, 600mAhrs cross- over point	12
Exponential, 200mAhrs cross-over point (a=1)	10
Exponential, 200mAhrs cross-over point (a=5)	14
Exponential, 600mAhrs cross-over point (a=1)	20
Exponential, 600mAhrs cross-over point (a=5)	21

Table 6.2 Network lifetime in the MPR/MBL scheme

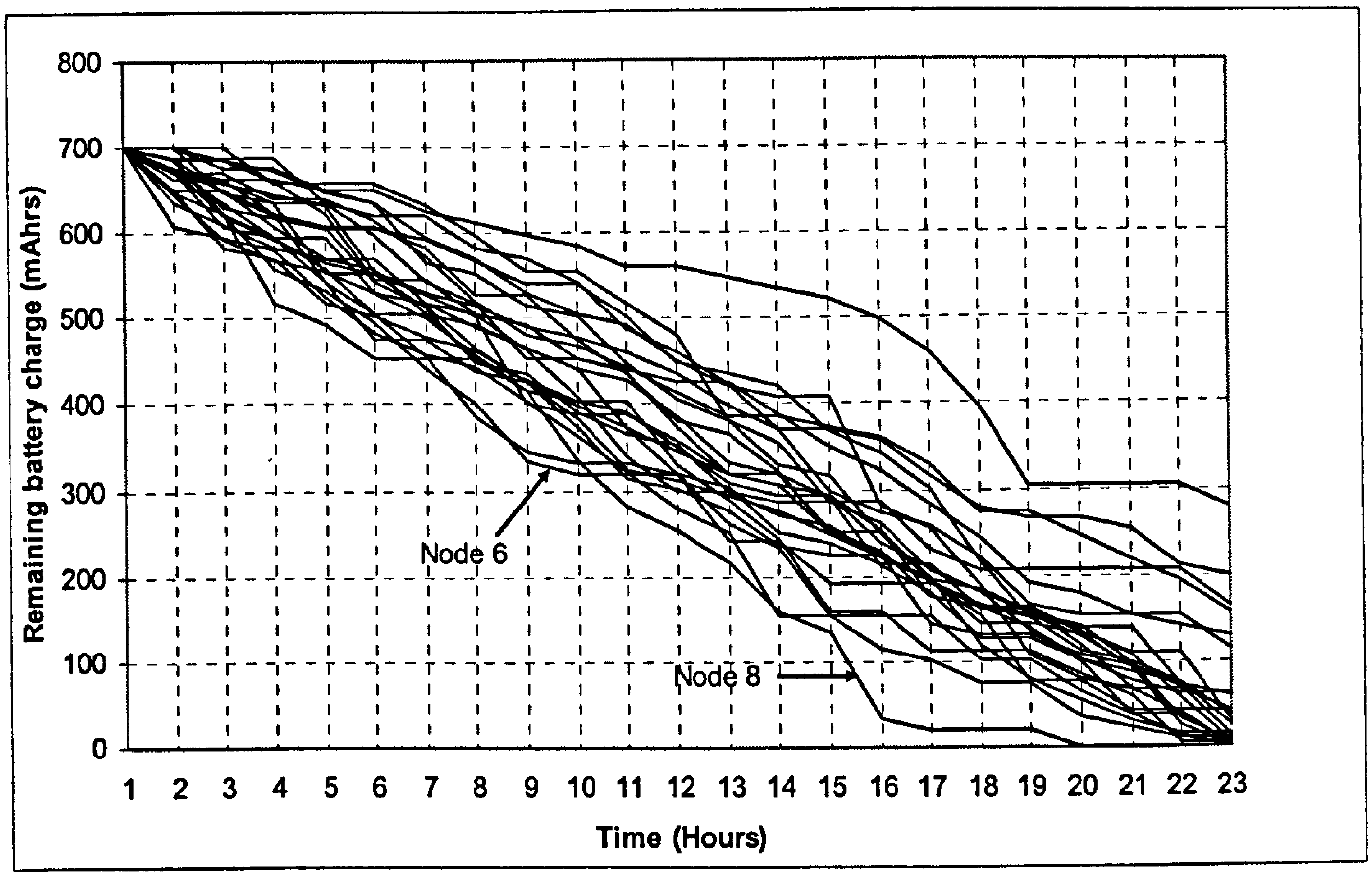


Figure 6.12 Battery charge vs time – MPR/MBR approach (cross over point 600mAhrs, a=1)

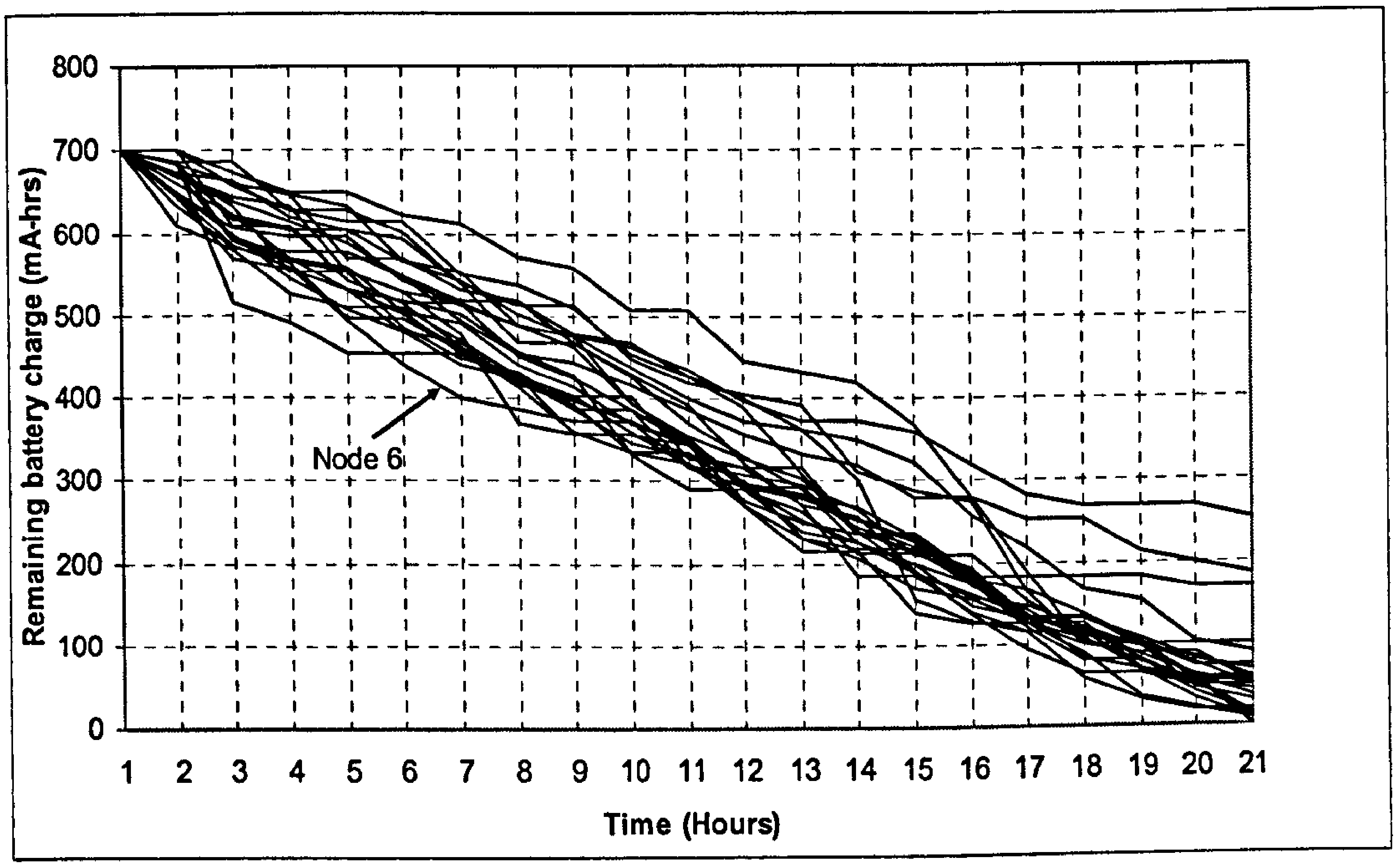


Figure 6.13 Battery charge vs time – MPR/MBR approach (cross over point 600mAhrs, a=5)

6.9 Discussion – network lifetime

6.9.1 Battery charge threshold approach

From the results shown in figure 6.9, it can be seen that the network lifetime depends on the threshold level and that there are two constraints. If a low threshold is selected, then heavily used nodes reach zero threshold before all nodes have reached the threshold level and so the battery usage is not evenly distributed between nodes. If the threshold level is set too high, then nodes reach threshold more rapidly and this can lead to route blockage. Nodes then need to be re-instated as relays and this leads to early exhaust of the more heavily used nodes. The maximum network lifetime is achieved with an intermediate threshold setting to give a balance between these two effects.

6.9.2 Residual battery charge routing

The residual battery charge approach selects a route to the destination on the basis of the minimum end to end sum of the inverse of the residual battery charge. The best route is the directly connected, however due to transmit power limitation of 100mW, the route is selected with the lowest number of hops. This approach does not take account of power consumption in the route selection and therefore source and relay nodes may be using more power than for the route that would be selected using the minimum power routing and MPR/MBL approach.

6.9.3 MPR/MBL approach

MPR/MBL approach using a linear relationship between cost and residual battery charge will not utilize the nodes batteries evenly. Using an inverse relationship as defined in equation 4.12, increases the residual battery charge cost when the available battery is low. This cost can be increased by using a higher order exponent. Hence it is possible to avoid using low battery nodes as relays. When a low cross over point is selected, then nodes continue to be used as relays until the residual battery charge reduces to a cross over point, and below that nodes with the lowest residual battery charge are not selected. If a cross over point is low, then the nodes only have a small amount of battery charge remaining for use as a source or a destination node and therefore rapidly depleted. However when a higher cross over point is used, then low battery nodes are avoided as relay nodes at an earlier stage, hence enough battery is left to use as a source or a destination node for a longer period of time.

6.9.4 Factors affecting network lifetime

An investigation has been carried out to identify parameters which are responsible for determining the network lifetime. Five different scenarios have been considered, and the description of each scenario is defined in table 6.3.

	Area	Processing Power	Data rate	Clutter factor range
Scenario A	50m x 50m	723 mW	6 Mbps	1-50 dB
Scenario B	100m x 100m	723 mW	6 Mbps	1-50 dB
Scenario C	50m x 50m	400 mW	6 Mbps	1-50dB
Scenario D	50m x 50m	723 mW	13 Kbps	1-50dB
Scenario E	50m x 50m	723 mW	6 Mbps	1-30 dB

Table 6.3 Scenarios description

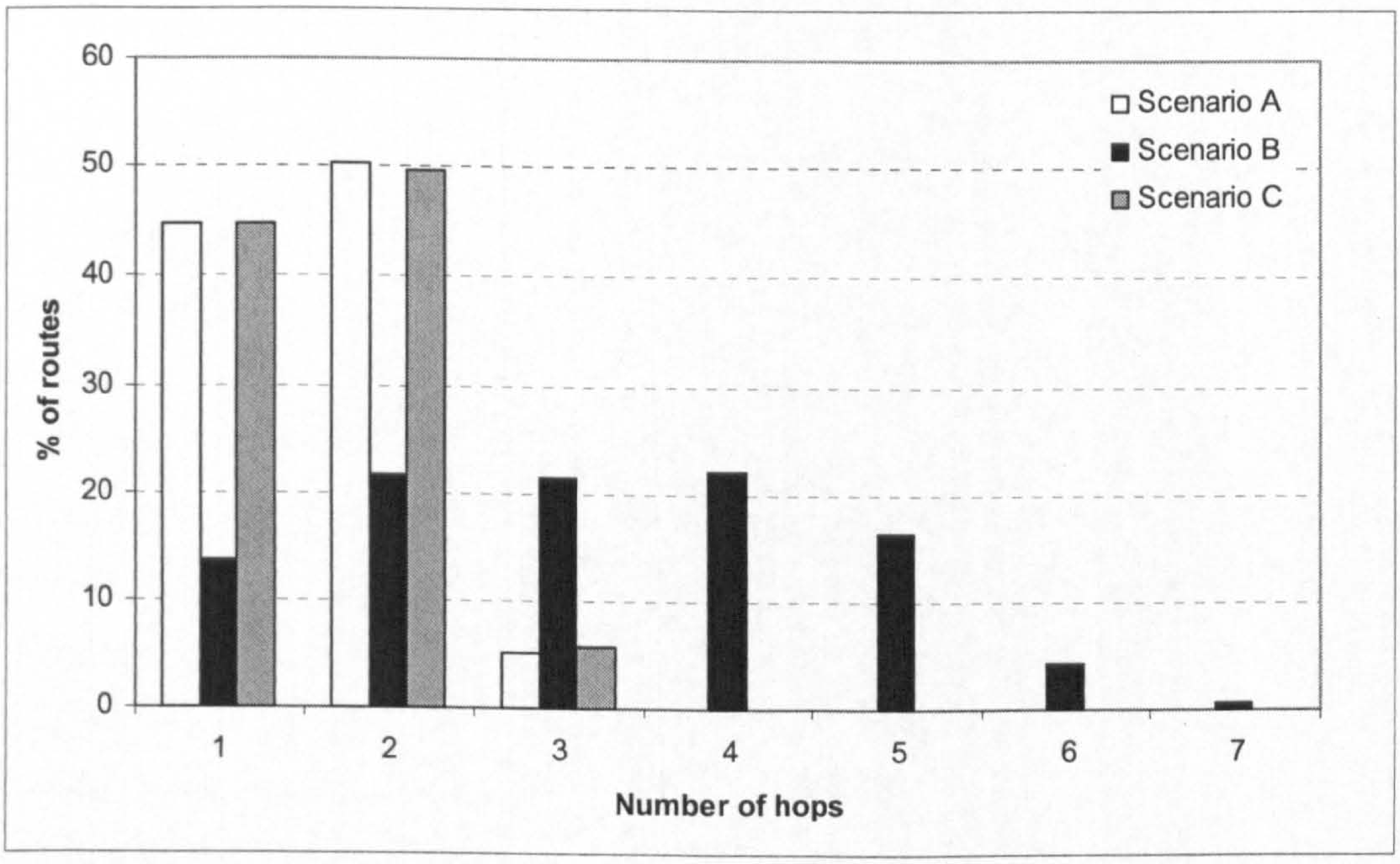


Figure 6.14 Number of hops in different scenarios

Scenario A is a reference scenario which is used throughout in the simulation. Figure 6.14 shows that the destination nodes are reachable up to 3 hops. Figure 6.15 shows the network lifetime of various approaches. In scenario A, the minimum power routing gives the lowest lifetime due to the fact that routing protocol does not take account of the residual battery charge in the route selection. Threshold scheme takes account of the residual battery charge in the route selection and power; therefore the network lifetime is extended. Residual battery charge routing takes only battery charge in the route selection, however the selected route may not be the minimum power route, hence there is a reduction in the network lifetime as compared to the MPR/MBL approach. MPR/MBL with inverse relationship and higher exponent order gives the maximum network lifetime compared to other approaches.

In scenario B, the nodes are distributed over a larger area (100m x 100m) with a random clutter factor in the range 1-50dB. As compared to scenario A, the lifetime is reduced from 9 to 5 hours with MPR. This is due to the fact that the transmission power increases because of the increase in the network area. Figure 6.14 shows that due to the larger network area (100m x 100m), some destination nodes are reachable up to 9-hops because of the transmit power limitation of 100mW. Figure 6.15 further shows that the use of a battery threshold gives a slight improvement in the network lifetime. This is because overused nodes are prevented from being used as relays.

Nodes in the scenario C use a lower processing power (400mW). Figure 6.15 shows that the network lifetime is 24 hours, and with a battery charge threshold of 30%, the network lifetime is increased to 34 hours. There is not significant difference in the number of hops in this scenario as compared to scenario A (see figure 6.14). If the processing power is less than the transmission power (<100mW) then the number of hops increases because the transmission power is more dominant. The transmission power is reduced if data is transmitted through one or more relay nodes [93]. This is already discussed in section 4.2, chapter 4. The residual battery charge and MPR/MBL schemes gives the same network lifetime, however further investigation reveals that residual battery charge scheme consumes more network power than MPR/MBL. This is discussed in the next section 6.10.

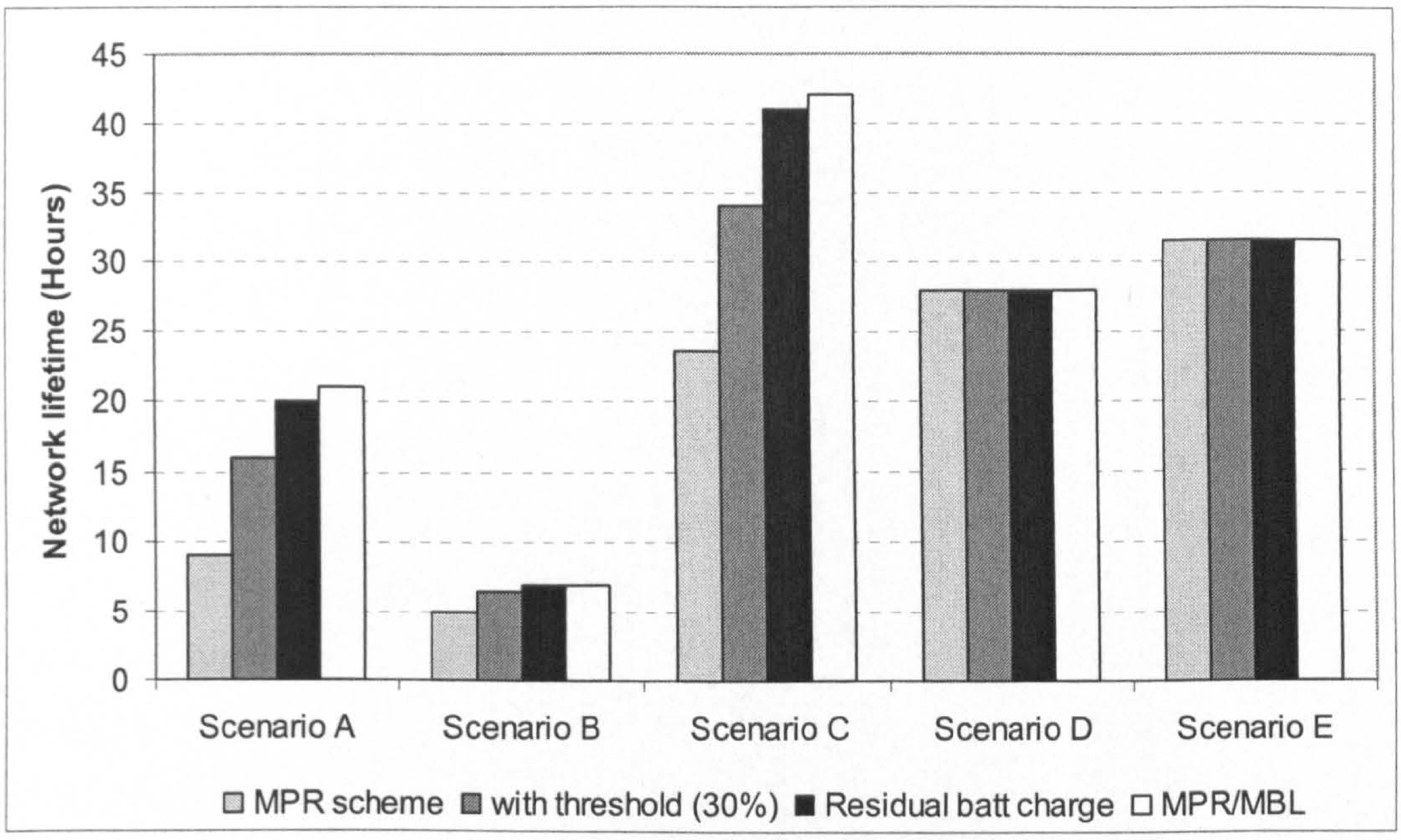


Figure 6.15 Effect on network lifetime in various scenarios

In scenario D, a lower bit rate of 13 kbps is selected which corresponds to that used to transport a voice service in the Groupe Spécial Mobile (GSM) network [108]. In this case the maximum path loss for each radio link carrying this service is larger than that for 6Mbps service used in scenarios A, B, C and E. As a result, approximately all of the links are directly connected. Figure 6.15 shows that the network lifetime is about 28 hours. However, there is no improvement in the network lifetime when other routing approaches are used. This is due to the reason that all of the minimum power routes are directly connected (1-hop) to the destination nodes. The required transmission power of all the source nodes to the destination node is under 100mW. If a relay is used, then this will add receive power (710mW) and processing power due to transmission (723mW) in addition to the transmission power. Thus the overall end to end power is much higher than the power consumed for the 1-hop link.

In scenario E, a lower attenuation range between 1-30dB is assumed between the nodes. This means that a larger range is possible for each radio link. Figure 6.15 shows that the network lifetime is 31 hours. Because all of the links are directly connected (similar to the scenario D), therefore there is no improvement in the network lifetime when other routing approaches are used.

The network lifetime shown in figure 6.15 is for the network topology shown in figure 6.2 in which nodes are fairly evenly distributed across the network area. Further tests have been carried out with different attenuation between the links with different node distributions. Network topologies and the network lifetime using various routing schemes are shown in Appendix N.

6.10 Power consumption

Figure 6.16 shows the network power consumption of different routing approaches as a function of time. Minimum power routing consumes the lowest power but has the lowest network lifetime. The MPR scheme does not utilize the nodes batteries evenly; therefore the most used node reach zero battery charge in 9 hours (see figure 6.6). The MPR approach is able to reduce power consumption where it is possible to use relay nodes to circumnavigate regions of high attenuation that would require high transmission power. The opportunities for reduced power route depends on the processing power in each relay node. The higher the processing power, the smaller the number of alternative routes with lower power consumption.

Battery charge threshold scheme uses minimum power routing therefore the power consumption is same as in MPR approach. In addition, the nodes which have residual battery charge are below the threshold are not considered as relays during minimum power route selection, therefore the network lifetime is increased (see figure 6.15). However, the power consumption increases with the increase in the number of isolated nodes. The routing scheme avoids low battery nodes which increases number of hops in the route.

The residual battery charge scheme selects a route on the basis of the residual battery charge. This approach provides even utilization of batteries however the scheme does not select lowest power route. As a result, the power consumption is increased.

Power aware routing discussed in section 4.3.2, chapter 4 is also analysed. The routing scheme takes account of transmission power and remaining battery capacity of the nodes. Figure 6.16 shows that the power consumption is same initially as compared to the minimum power scheme. However in an early stage, the network power consumption increases. This is due to the fact batteries of the nodes are used in the minimum power route, and the battery term in equation 4.4 becomes dominant. As a result, alternate route is selected which consumes higher power than the minimum power route.

MPR/MBL scheme gives the longest network lifetime. The network power consumption of MPR/MBL follows MPR initially as seen in figure 6.10. This is valid because MPR/MBL uses MPR scheme initially when the nodes have full residual battery charge. However, with the time, the network power increases. This is due to the fact that most used nodes are avoided in the route. The alternative route is selected which consumes higher power than the minimum power route.

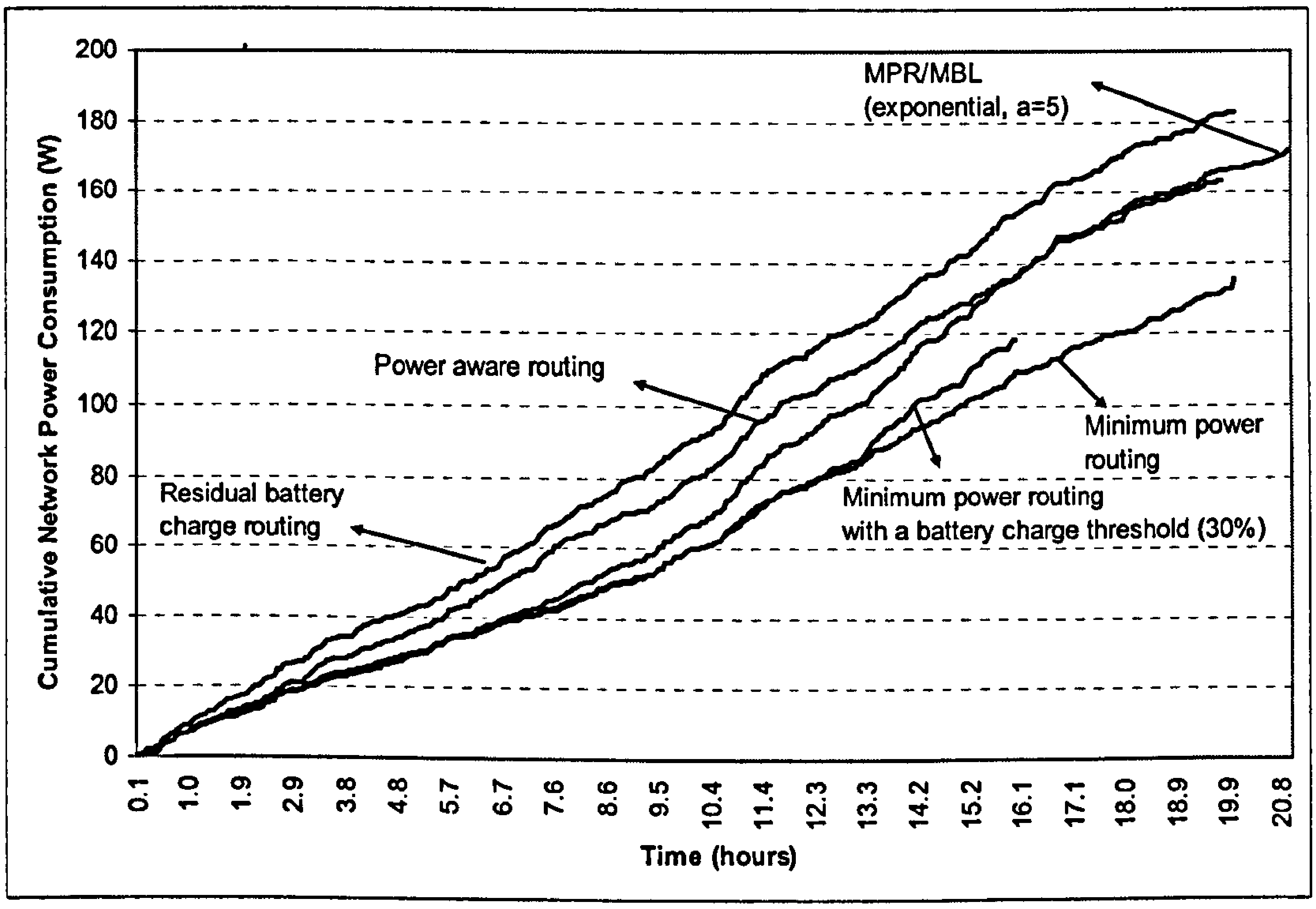


Figure 6.16 Network power consumption between various routing schemes

6.11 Summary

In this chapter, simulation model has been described and the network lifetime has been investigated for various routing approaches. Table 6.4 shows the comparison of the network lifetime.

Routing schemes	**Network lifetime (Hrs.) (without throughput maintained)**
Minimum power	9
Minimum power with a battery charge threshold (30%)	16
Residual battery charge	20
Power aware routing	20
MPR/MBL (exponential, cross-over point 600mAhrs, a=5)	21

Table 6.4 Comparison of the network lifetime between various routing schemes

When minimum power routing is used, then some of the relay nodes are heavily used and therefore the batteries became depleted more rapidly than other nodes. The network lifetime when the first node reached end of battery charge is 9 hours. In order to prevent nodes from being overused, a threshold scheme is used in which nodes are available as relays up to a specified threshold. Beyond this threshold, they are withdrawn from use as relays. The minimum power route is selected using available relay nodes. Network lifetime depends upon the value of the threshold. A higher threshold provides route blockages at an early stage, and a lower threshold can lead to some nodes being overused. The optimum threshold has been found to be 30% and this provides the network lifetime of about 16 hours. Another approach is used which selects routes on the basis of available battery of the nodes. This scheme utilizes batteries of the nodes evenly however the overall power consumption increases. The network lifetime is 20 hours. Power aware routing selects routes on the basis of power and available battery of the nodes. This scheme gives the network lifetime of 20 hours.

A MPR/MBL scheme is proposed which aims to minimise the power consumption and maximise the network lifetime. Minimum power routes are selected when all nodes have adequate battery charge however when node's battery charge is low then such nodes are avoided as relays. This approach gives the network lifetime of 21 hours which represents an improvement by 133% over the MPR scheme, 31% over the MPR with battery charge threshold scheme, and 5% over the power aware routing scheme and the residual battery charge scheme.

7. Enhanced System Design

7.1 Introduction

This chapter discusses enhancements to the system design discussed in chapter 5. Firstly techniques for supporting multiple simultaneous access are investigated. Secondly techniques for making the throughput per unit time independent of the number of hops in a multi-hop route are analysed.

7.2 Multiple access approaches in multi hop ad hoc network

7.2.1 Spatial re-use in space division multiple access (SDMA) approach

If a single frequency is used in an ad hoc network, then the ability to establish other routes at the same time will be determined by interference. The original 25 node model described in chapter 6 using the same simulation parameters defined in table 6.1 has been extended to analyse spatial re-use in a single frequency network.

Figure 7.1 shows an illustration of simultaneous routes in the network. It is assumed that the base route (nodes 13-22) is active at the start of the network. Node 13 is using minimum transmission power which is calculated with the help of path loss and receiver sensitivity, see section 6.2, chapter 6. This path loss is valid when there is no interference (such as the case when only one route exists in the network). If a second route (nodes 1-2) is required after the route 13-22 has been established, then there are two conditions that need to be met:

i) the maximum transmission power for node 1 is determined by the interference it will cause to node 22, and

ii) the minimum transmission power for node 1 is determined by the power needed to give adequate signal to interference noise (SINR) at node 2 taking account of the transmission power from node 13.

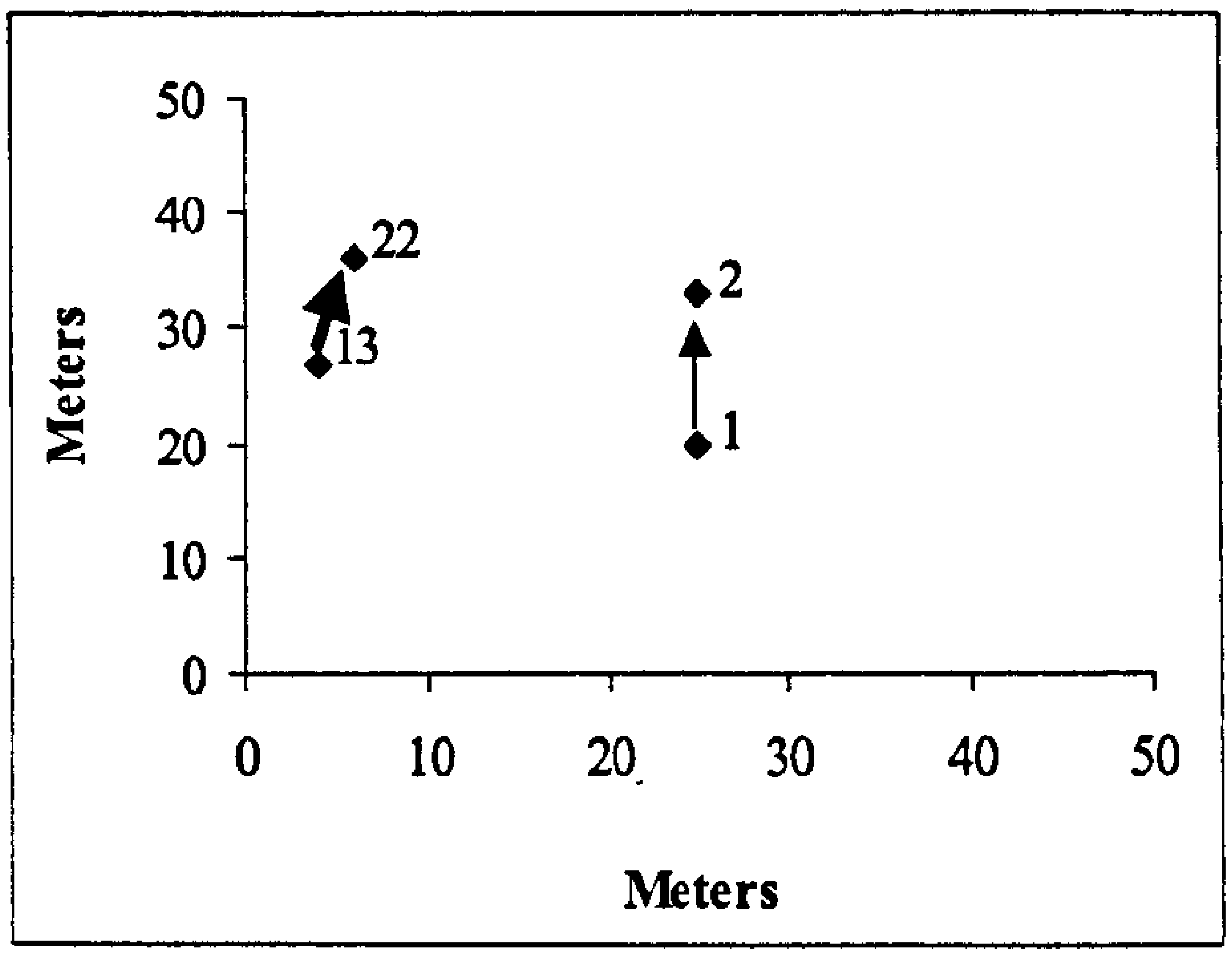


Figure 7.1 Illustration of simultaneous routes

SDMA approach is investigated in a 25-node network model shown in figure 7.2 to identify simultaneous routes in the network after a base route (nodes 13-22) has established. The source and the destination routes are selected sequentially. It is assumed that if one node becomes a source, relay or destination node for the route, then that node is reserved for the route, and will not take part in any other routes during the length of the reservation. The first route in the sequence is (nodes 1-2). The two conditions (discussed in section 7.2.1) are met to establish a simultaneous route (nodes 1-2) along with the base route (nodes 13-22). However the route (nodes 1-2) creates interference to other nodes in the network. The nodes 1, 2, 13 and 22 are already engaged in the communication; therefore the next route in the sequence is (nodes 3-4). The route (nodes 3-4) is a multi hop route. The nodes 3 and 15 can only transmit if their transmission powers do not create interference to the neighbouring receivers which are nodes 22 and 2. The minimum transmit power levels at nodes 3 and 15 are determined by the power needed to give adequate signal to interference noise (SINR) at nodes 15 and 4 respectively taking account of the transmission power from nodes 13 and 1. The conditions for the route (nodes 3-4) are met, and the route (nodes 3-4) can be established simultaneously. However the route (nodes 3-4) also creates interference to other nodes.

The routes are scanned sequentially and identify simultaneous routes based on the methodology described above. Figure 7.2 shows the node distribution in the network, and figure 7.3 shows the random attenuation between the nodes. Figure 7.2 shows that a total of 8 simultaneous routes (including two multi hop routes) can co-exist with the single hop base route (13-22). Some of the nodes are not able to establish simultaneous routes such as nodes 16, 17, 18, 21 and 23. This is due to the fact that the nodes in the routes create unacceptable interference to the other routes.

Simultaneous routes are calculated with the help of the allowed transmitter power and the minimum received power. The calculations are described with the help of a 6-node example in Appendix J.

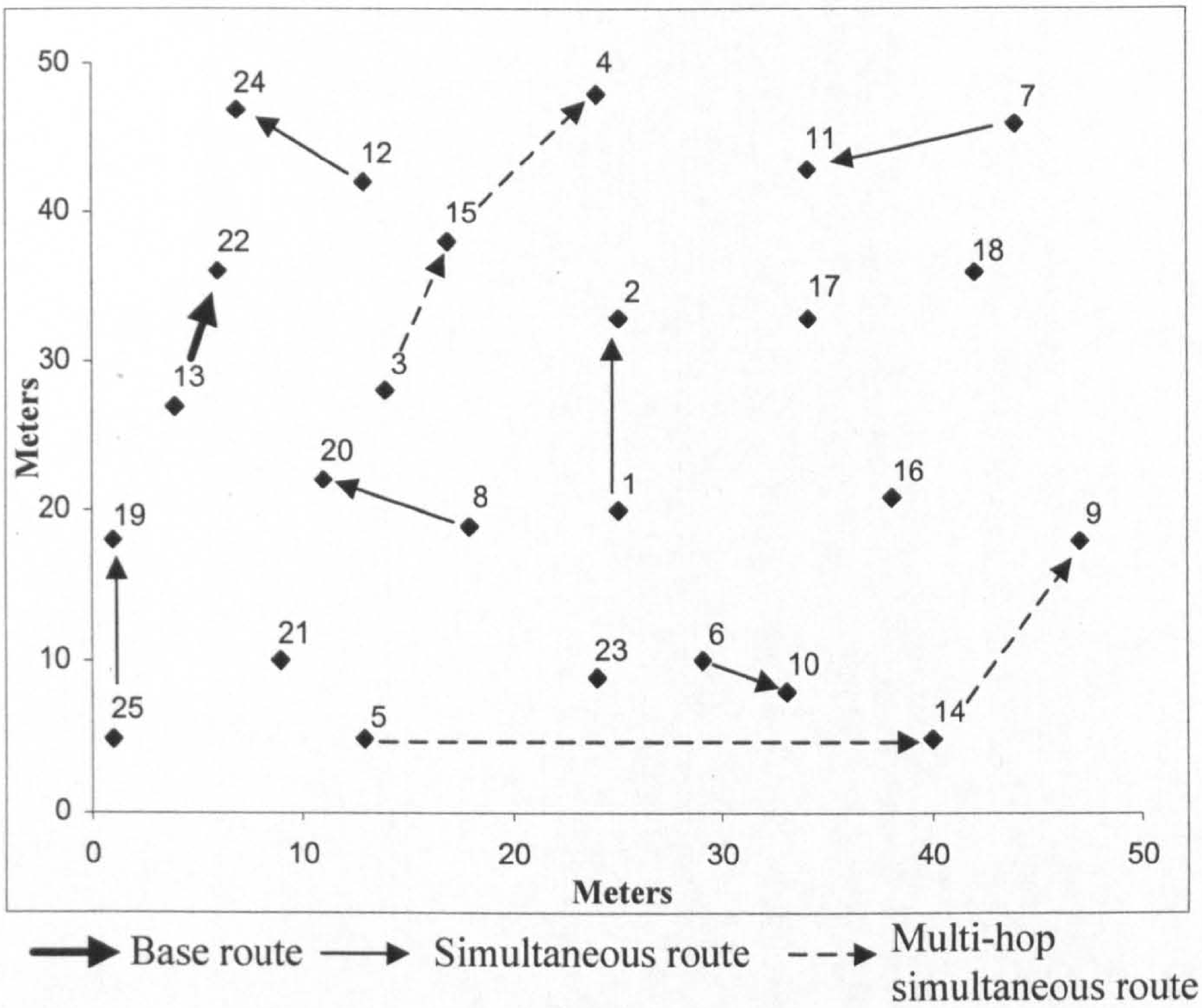


Figure 7.2 Simultaneous routes in the network model – base route (13-22)

Attenuation in dB

Nodes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1		48.5	47.3	47.6	44.6	44.8	48.9	31.5	48.8	49.1	48.5	36.6	46.2	49.4	49.0	45.7	49.8	49.4	37.5	49.8	45.6	47.2	48.8	37.3	47.7
2	48.5		46.7	44.7	47.9	48.1	44.2	44.5	49.2	49.2	47.7	49.9	49.6	43.6	48.4	49.9	43.9	47.3	40.3	50.0	48.3	32.0	47.6	40.0	40.1
3	47.3	46.7		49.0	44.5	36.6	44.7	45.8	44.8	49.8	49.9	46.0	44.4	42.1	42.1	48.1	46.1	46.2	48.5	45.1	48.0	43.2	42.7	47.7	39.1
4	47.6	44.7	49.0		46.6	49.6	44.2	48.9	45.8	44.6	49.6	48.0	48.0	46.3	39.9	47.5	48.4	49.6	49.2	33.5	47.4	49.6	46.3	48.3	47.0
5	44.6	47.9	44.5	46.6		47.1	46.7	45.5	46.1	44.3	37.5	43.9	49.9	37.8	45.9	45.6	46.9	41.9	46.8	44.1	48.0	47.3	41.9	49.7	48.2
6	44.8	48.1	36.6	49.6	47.1		47.0	45.9	40.3	48.9	46.6	48.8	47.8	49.2	32.7	43.2	38.7	40.2	45.2	41.1	14.0	47.3	48.2	47.4	49.2
7	48.9	44.2	44.7	44.2	46.7	47.0		39.1	42.8	48.3	46.6	45.5	41.8	48.5	49.7	47.2	39.5	48.8	46.0	46.6	46.9	43.2	45.2	39.8	47.7
8	31.5	44.5	45.8	48.9	45.5	45.9	39.1		42.3	49.7	39.9	46.5	44.4	49.4	48.8	44.4	48.3	44.1	39.5	34.9	45.1	49.0	44.7	43.7	46.8
9	48.8	49.2	44.8	45.8	46.1	40.3	42.8	42.3		44.1	45.3	36.5	46.8	43.1	49.4	47.7	48.8	49.7	45.2	47.3	39.1	48.0	46.1	49.8	40.6
10	49.1	49.2	49.8	44.6	44.3	48.9	48.3	49.7	44.1		49.7	47.9	45.4	41.7	46.8	43.4	50.0	41.2	34.6	45.4	47.4	49.7	47.3	46.1	49.3
11	48.5	47.7	49.9	49.6	37.5	46.6	46.6	39.9	45.3	49.7		49.2	48.3	48.6	50.0	45.3	46.9	46.2	48.4	42.5	46.3	47.3	49.1	47.3	46.3
12	36.6	49.9	46.0	48.0	43.9	48.8	45.5	46.5	36.5	47.9	49.2		47.1	43.6	47.9	46.9	48.3	49.5	45.7	44.8	44.7	41.8	47.2	43.5	47.7
13	46.2	49.6	44.4	48.0	49.9	47.8	41.8	44.4	46.8	45.4	48.3	47.1		45.6	49.4	46.8	42.8	48.4	48.7	47.9	48.9	42.1	49.1	43.1	49.8
14	49.4	43.6	42.1	46.3	37.8	49.2	48.5	49.4	43.1	41.7	48.6	43.6	45.6		38.2	37.9	49.0	45.8	46.7	40.8	40.6	42.4	36.8	48.5	47.3
15	49.0	48.4	42.1	39.9	45.9	32.7	49.7	48.8	49.4	46.8	50.0	47.9	49.4	38.2		47.5	43.4	46.7	48.7	48.8	46.0	49.6	48.7	39.5	48.0
16	45.7	49.9	48.1	47.5	45.6	43.2	47.2	44.4	47.7	43.4	45.3	46.9	46.8	37.9	47.5		48.5	32.0	46.3	46.0	44.4	49.9	49.0	48.4	46.2
17	49.8	43.9	46.1	48.4	46.9	38.7	39.5	48.3	48.8	50.0	46.9	48.3	42.8	49.0	43.4	48.5		48.7	44.4	45.5	46.4	49.8	40.8	48.1	45.4
18	49.4	47.3	46.2	49.6	41.9	40.2	48.8	44.1	49.7	41.2	46.2	49.5	48.4	45.8	46.7	32.0	48.7		40.2	42.7	38.9	46.4	49.8	47.3	46.9
19	37.5	40.3	48.5	49.2	46.8	45.2	46.0	39.5	45.2	34.6	48.4	45.7	48.7	46.7	48.7	46.3	44.4	40.2		49.9	43.4	45.8	46.0	44.5	47.0
20	49.8	50.0	45.1	33.5	44.1	41.1	46.6	34.9	47.3	45.4	42.5	44.8	47.9	40.8	48.8	46.0	45.5	42.7	49.9		41.4	47.1	49.8	47.5	49.6
21	45.6	48.3	48.0	47.4	48.0	14.0	46.9	45.1	39.1	47.4	46.3	44.7	48.9	40.6	46.0	44.4	46.4	38.9	43.4	41.4		48.2	46.4	48.4	38.1
22	47.2	32.0	43.2	49.6	47.3	47.3	43.2	49.0	48.0	49.7	47.3	41.8	42.1	42.4	49.6	49.9	49.8	46.4	45.8	47.1	48.2		48.8	48.5	47.0
23	48.8	47.6	42.7	46.3	41.9	48.2	45.2	44.7	46.1	47.3	49.1	47.2	49.1	36.8	48.7	49.0	40.8	49.8	46.0	49.8	46.4	48.8		41.9	43.5
24	37.3	40.0	47.7	48.3	49.7	47.4	39.8	43.7	49.8	46.1	47.3	43.5	43.1	48.5	39.5	48.4	48.1	47.3	44.5	47.5	48.4	48.5	41.9		45.1
25	47.7	40.1	39.1	47.0	48.2	49.2	47.7	46.8	40.6	49.3	46.3	47.7	49.8	47.3	48.0	46.2	45.4	46.9	47.0	49.6	38.1	47.0	43.5	45.1	

Figure 7.3 Random attenuation in dB between the nodes

Further investigation has been carried out to understand the effect on simultaneous routes when a multi hop route is selected as the base route. Figure 7.4 shows a multi hop route (nodes 7-17-3-25). With this base route only 4 other simultaneous routes can co-exist in the network.

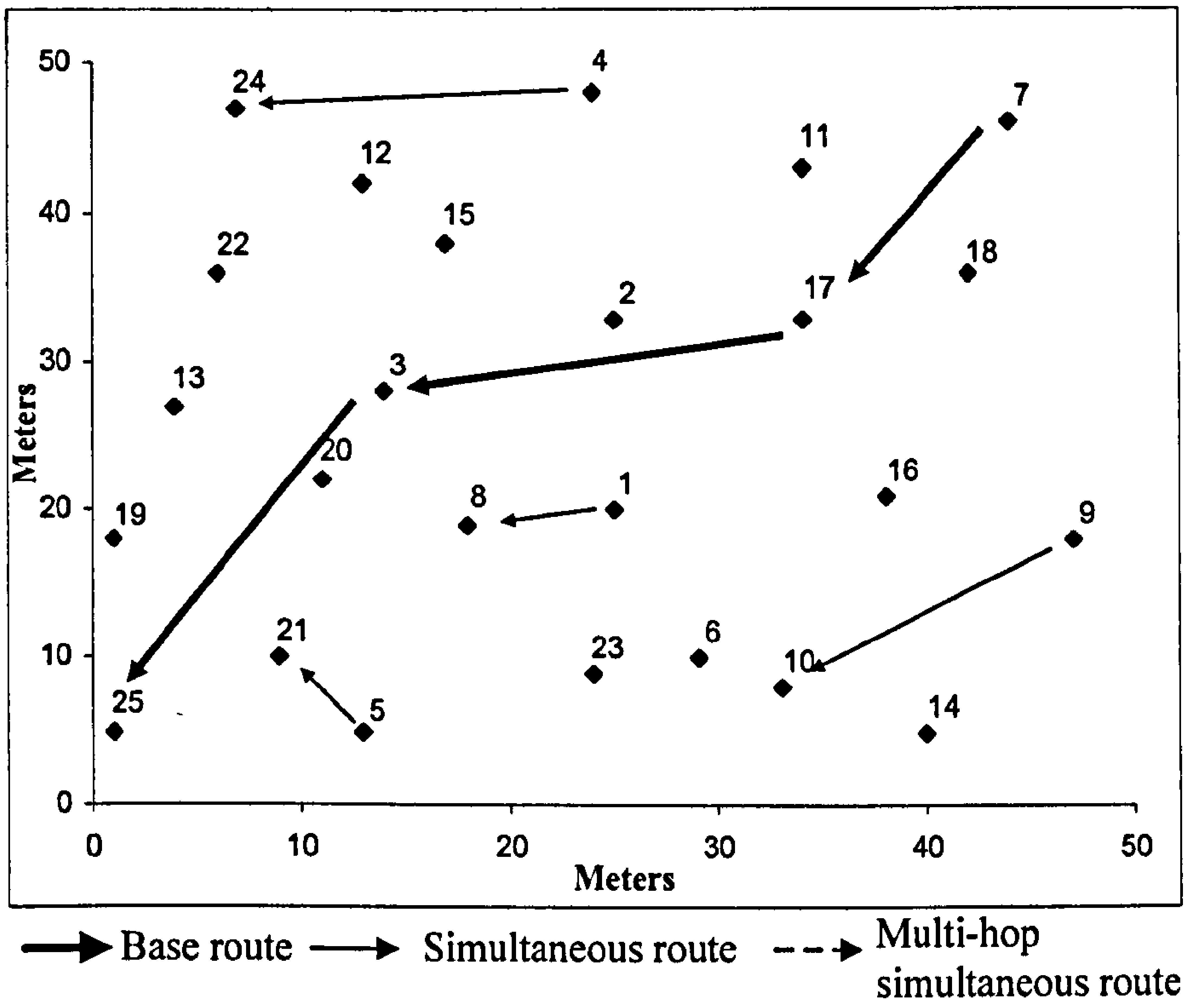


Figure 7.4 Simultaneous routes in the network model – base route (7-17-3-25)

The results gathered from the single-hop and multi-hop base routes are valid only if the base route shown in figure 7.2 and figure 7.4 is used, source and destination nodes are selected sequentially, using the same topology and attenuation between the nodes. However, the number of simultaneous routes in the network depends upon the selection of the base route, the order the source and the destination nodes are selected, topology and attenuation between the nodes.

This investigation shows that the SDMA approach does not provide guaranteed access in the network.

7.2.2 Spatial re-use in FDMA approach

Simultaneous communication and guaranteed access between ad hoc nodes is important particularly when an ad hoc network is deployed as an emergency network in a disaster area. This problem has been discussed in the literature. Nasipuri et al. [113], Jain et al. [114], Wu et al. [115] and So et al. [116] propose a range of schemes in which nodes in a wireless network use different frequencies for simultaneous transmissions. All of these schemes refer to a single hop wireless links.

The IEEE 802.16j (Relay Task Group) is currently developing a draft for "Air Interface for Fixed and Mobile Broadband Wireless Access Systems – Multihop Relay Specification" [117]. In one of the contributed documents [118], researchers proposed a similar approach of using multiple channels in multi hop links as discussed in the literature above. Multi hop links can use the same carrier frequency per hop separated in time or different carrier frequencies per hop and may not be time separated. The former is termed as in-channel relay and latter is termed as out-of-channel relay [119].

In the next section, the multiple frequencies approach is analyzed using a network model of 25 nodes.

7.2.2.1 Spatial re-use with 3 frequency channels

Systems based on the IEEE 802.11b/g standards operating in the 2.4GHz Industrial, Scientific and Medical (ISM) band have 14 carrier frequencies in which 3 frequencies can be used simultaneously in the network [120]. In the 5GHz Unlicensed National Information Infrastructure (U-NII/ISM) band such as systems based on the IEEE 802.11a standard can use 23 carrier frequencies simultaneously [121]. If each route is allocated a single frequency, then depending on the IEEE 802.11 standard being used, up to 3 or 23 simultaneous routes can be established (assuming no frequency re-use).

Spatial re-use for the FDMA approach has been calculated for the 25 node network model using three carrier frequencies as defined in IEEE 802.11b/g. Figure 7.5 shows that 24 simultaneous routes can exist in the network. Frequencies 1, 2 and 3 represents channels 1, 6 and 11 in IEEE 802.11b/g. Further investigation shows that a node selected as a source, relay or destination (SRD) for one route using one frequency can be selected as SRD for another route using another frequency. This assumes that each node has multiple radios that can operate independently, rather than one radio that can be tuned. For example, in figure 7.5, node 3 is the source node for the route 3-15-4 using frequency 1, and the same node 3 is the destination node for the route 2-3 using frequency 2.

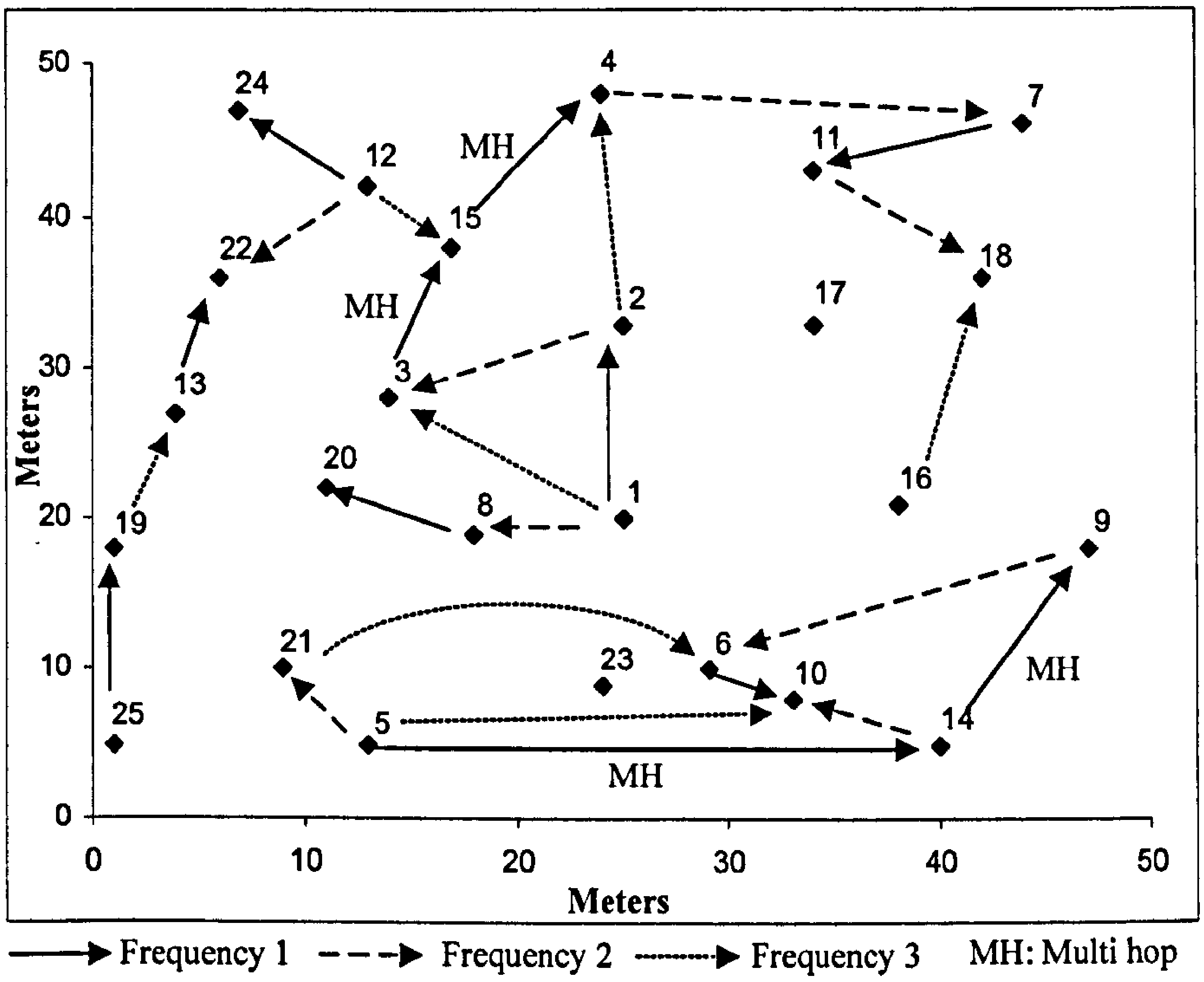


Figure 7.5 Simultaneous routes in the network model using three carrier frequencies

If guaranteed access is needed for a 25 node network, then each node is allocated a single frequency. If network uses a licensed frequency then multiple frequencies for data channel could be a costly solution. Moreover, the design of a transceiver, capable of tuning into multiple frequencies will be more complex and hence more costly.

7.2.3 Spatial re-use in TDMA approach

In this approach, a time slot is allocated to each node in the network. However, allocating different time slots to each user reduces the bit rate per user, or alternatively a higher bit rate is needed to maintain the bit rate per user.

7.3 Throughput in multi hop links

When a multi hop route is formed in a single frequency network, then it is not possible to use the same frequency for the transmit and receive side of a relay node because it is not possible to provide sufficient isolation between the transmitter and receiver. Therefore the transmission and reception must be separated in time. As a result, if a source node transmits 'x' packets in time period 'T' over a single hop route, then for a two hop route, as the relay node cannot transmit and receive at the same time, then for half the time period is allocated to the first hop and half for the second hop – hence the throughput per unit time is reduced by x/2. Similarly, for a three hop route, the throughput per unit time is reduced by 'x/3', [93]. In general, the throughput per unit time of a route is x/N where N is the number of hops in the routes, this is shown in figure 7.6.

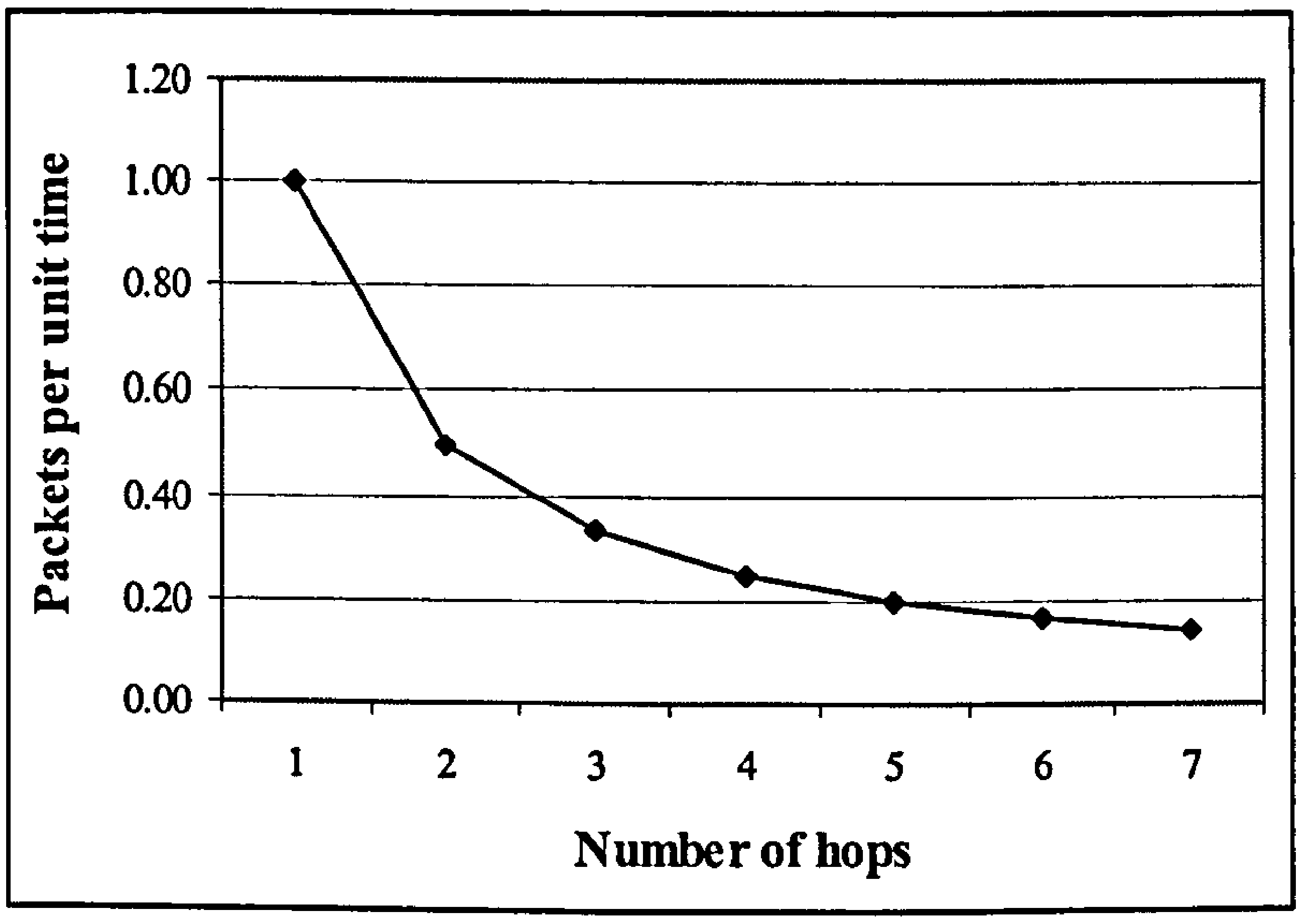


Figure 7.6 Reduction in packets per unit time with increase in number of hops

One method of maintaining the throughput is to use a higher order modulation scheme. The relative carrier to interference ratio (C/I) for four modulation schemes is given in table 7.1 [122]. Assuming that BPSK modulation can provide the required throughput (6 Mbps) for a single hop route, then the same capacity is possible over a two hop route using QPSK with a 3dB reduction in C/I margin; a three hop route will require 8QAM with a 6.7dB reduction in C/I margin, a four hop route will require 16QAM with a 10dB reduction, a 6 hop route will require 64QAM with a 16.2dB reduction in C/I margin.

Modulation	BPSK	QPSK	8QAM	16QAM	64QAM
C/I required relative to BPSK	0 dB	+3 dB	+6.7 dB	+10 dB	+16.2 dB

Table 7.1 C/I for different modulation schemes relative to BPSK

In order to maintain the C/I ratio, the transmitter power must be increased when higher order modulation is used. However an increase in the transmitter power also increases the interference footprint of the transmitter in the network which decreases the number of simultaneous routes. For example, figure 7.7 shows that five routes can be used simultaneously with the multi hop route (7-17-3) using BPSK modulation. However when the same two-hop route uses QPSK modulation to maintain the throughput per unit time, then the 3dB increase in transmission power required at nodes 7 and 17 decreases the number of simultaneous routes to four, see figure 7.8.

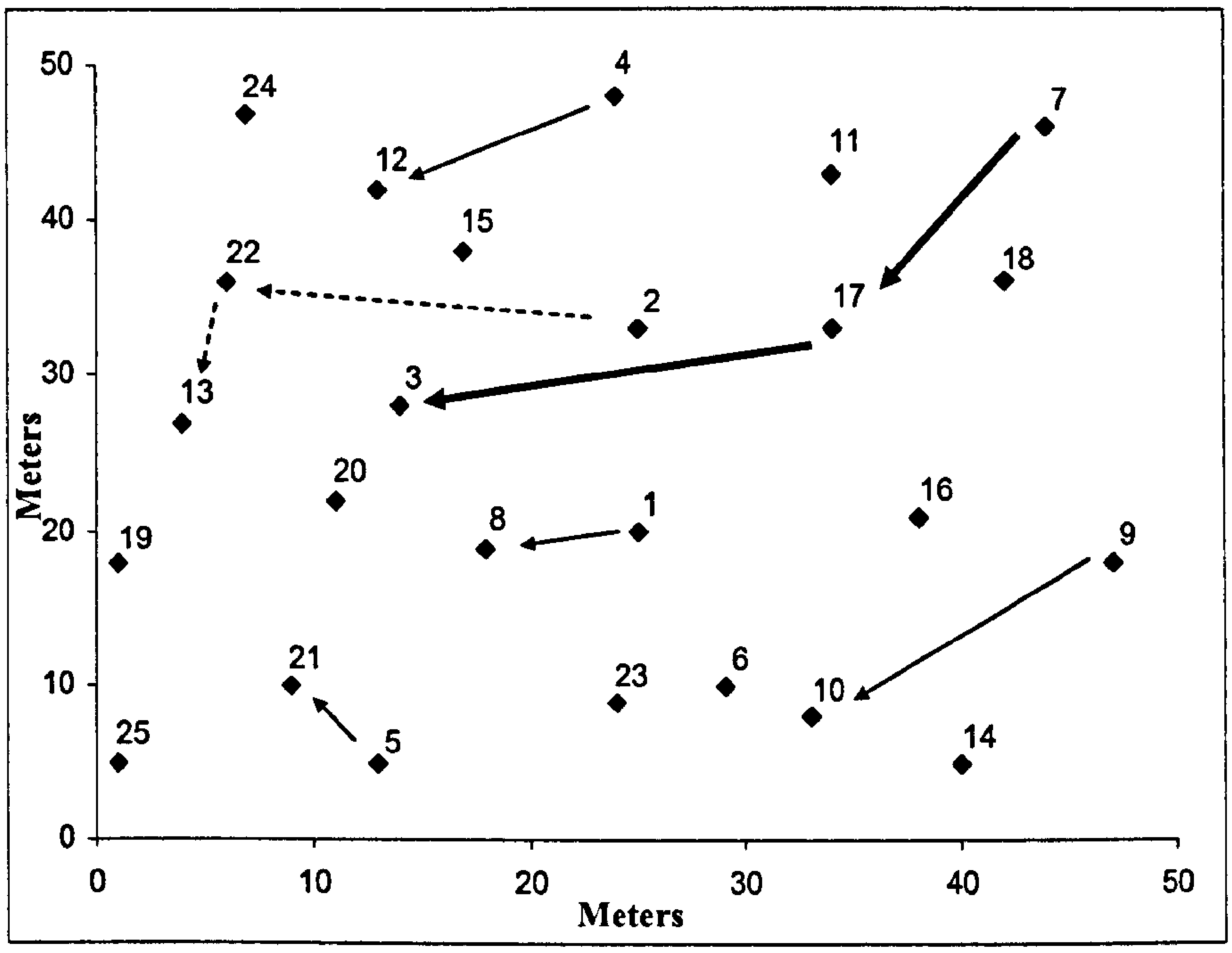


Figure 7.7 Multi hop route without adaptive modulation (BPSK)

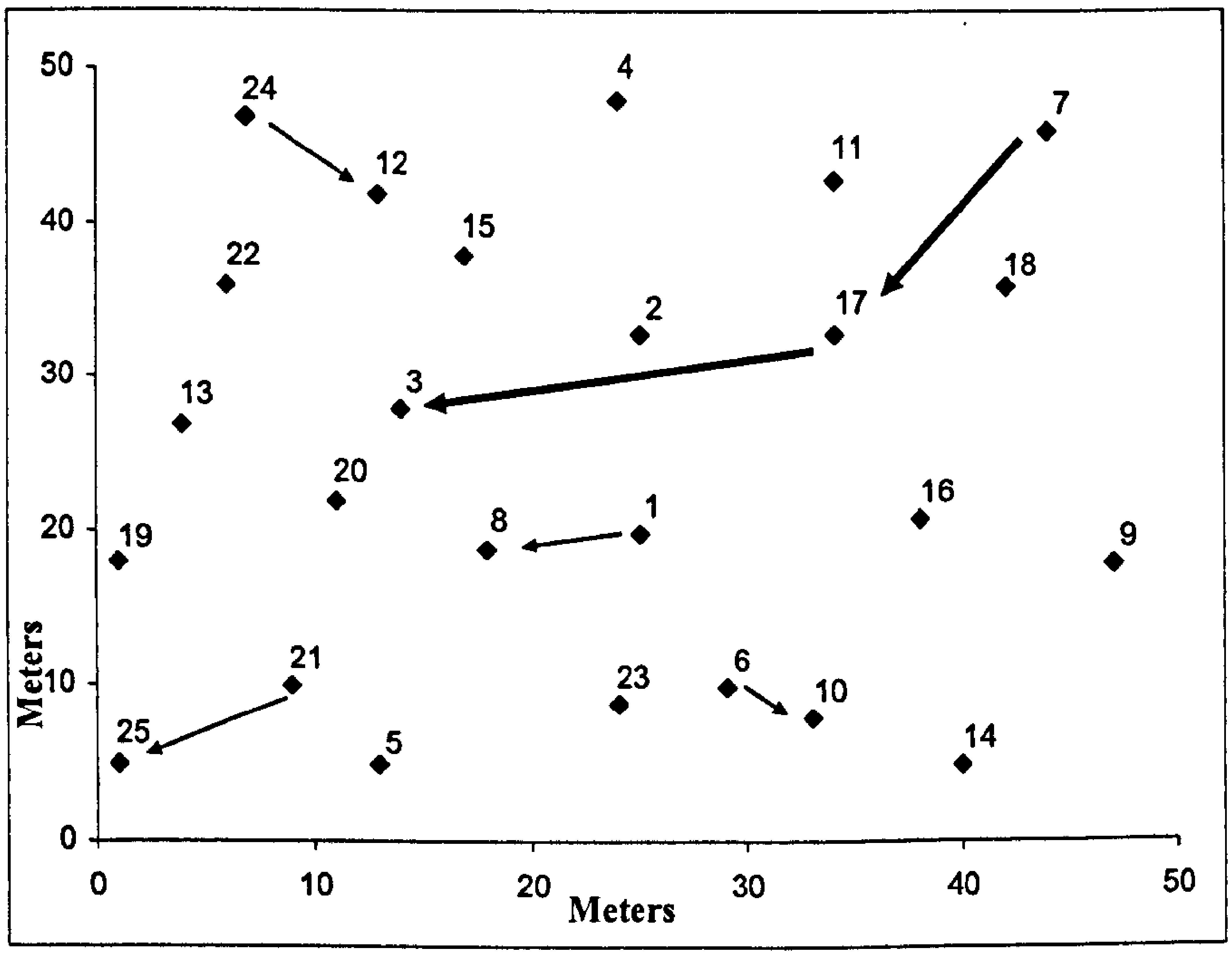


Figure 7.8 Multi hop route with adaptive modulation (QPSK)

It should be noted that it may not always be possible to establish routes using higher order modulation. For example, in figure 7.4, using BPSK modulation, node 17 requires +12dBm for the link to node 3. If the modulation is changed to 16QAM, then the transmitter power needs to be increased by 10dB to maintain the C/I ratio. The maximum transmitter power however is limited to +20dBm (see section 6.1), so it is not possible to maintain the C/I ratio with 16QAM modulation over this link.

7.4 Multiple access and throughput in the enhanced system design

The system design discussed in chapter 5 assumes that only one node transmits or receives data at one time. In addition, the design does not maintain throughput per unit time. However, in the enhanced system design, the TDMA scheme has been selected because of the limited access with spatial re-use and the cost and complexity of using FDMA scheme. Moreover adaptive modulation is used in the routes to maintain throughput per unit time.

7.4.1 Time slotted data channel

The "Time slot (data)" field is 8 bits long which gives 256 available time slots (can accommodate large number of nodes). Each node is allocated a time slot.

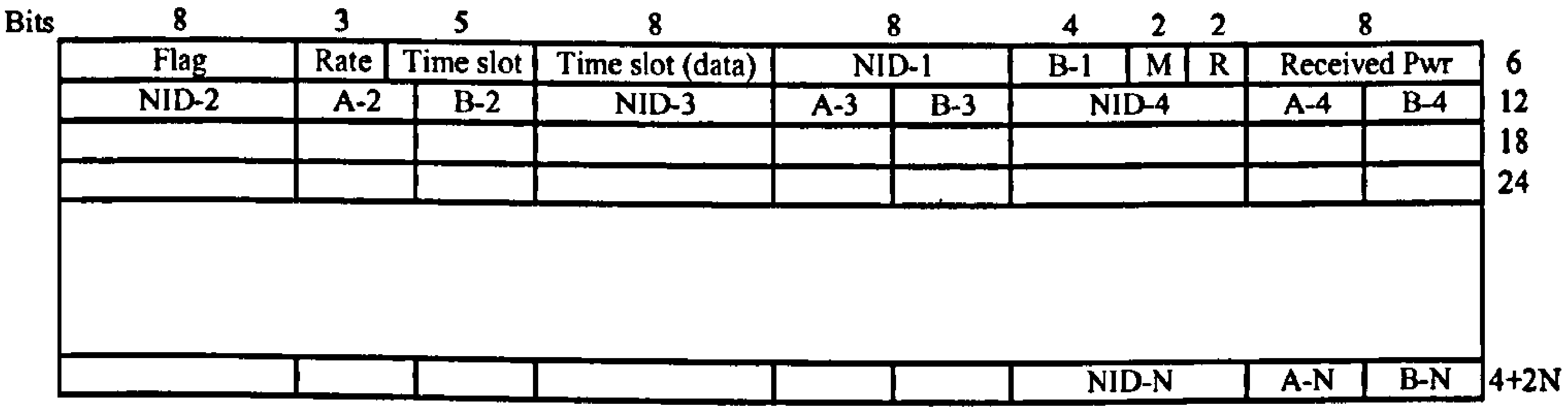


Figure 7.9 Enhanced beacon frame

7.4.2 Adaptive modulation in the enhanced system design

The source node selects the required modulation based on the number of hops. In addition, the source node also checks whether the additional transmitter power needed to maintain the received carrier to noise ratio is within the maximum transmit power limit (+20dBm). When the source node selects the route, it sends a RTS frame to reserve relay nodes in the route (as discussed in section 5.4.2). However, an extra field "Mod" is added to notify relay nodes about the selected modulation, see figure 7.10.

SA=source address, DA=destination address
RA-1=first relay address, RA-N=last relay address, FCS=frame check sequence
Mod = selected modulation, Res = reserved for future use

Figure 7.10 Notification of selected modulation through RTS frame

7.4.3 Cross layer interactions in the enhanced system design

In the system design, route selection at the network layer is achieved with the help of information from other layers. Figure 7.11 shows the interactions such as:

Physical-Network cross-layering: Attenuation and received power information are shared with the network layer. The attenuation information is required to calculate minimum transmission power for a route. The received power information is needed to know the maximum allowed transmission power.

Application-Network-Physical cross-layering: The application layer defines the required quality of service (throughput) and passes this to the network layer. The routing protocol searches for the lowest power route and checks if it is capable of using higher order modulation to maintain the throughput per unit time. When the route is found, then the network layer instructs the physical layer to use the chosen modulation. Battery information is made available to the network layer. The working of the enhanced system design using different routing schemes is shown in Appendix K, L and M.

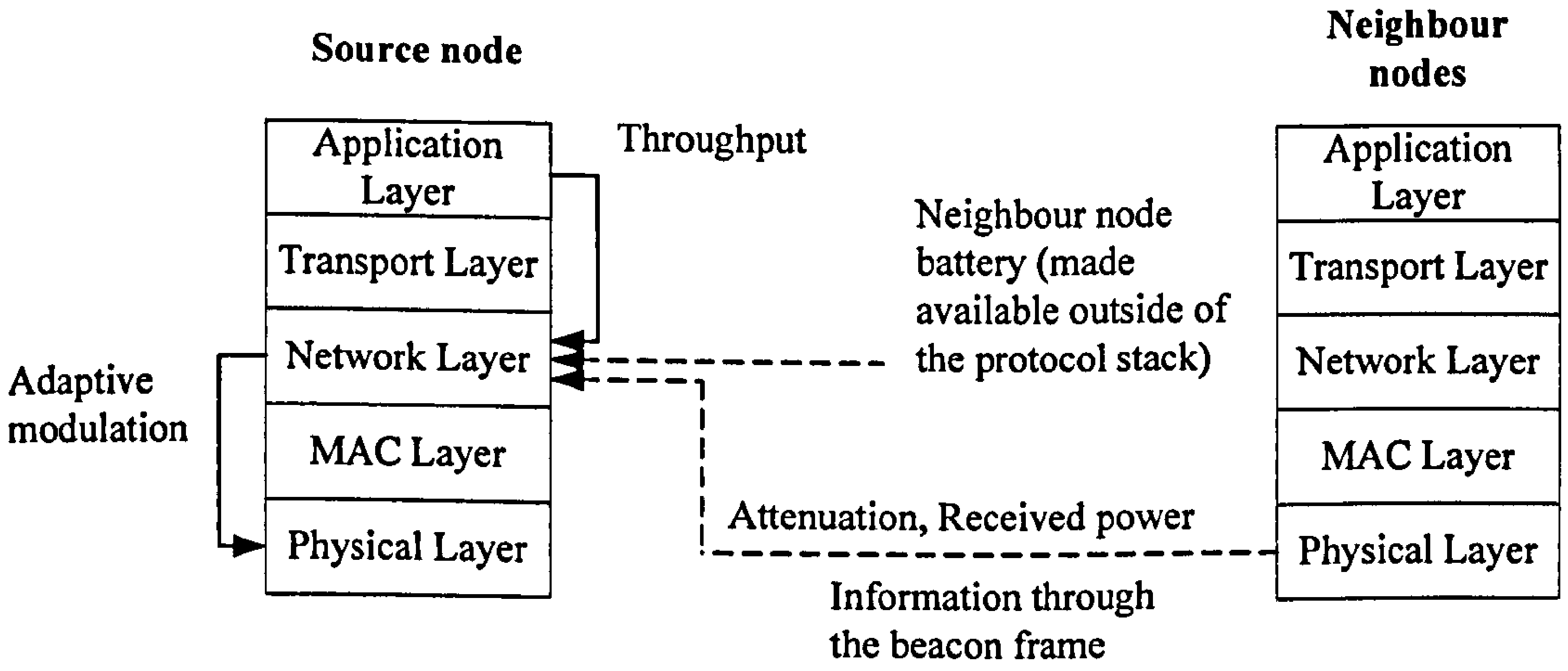


Figure 7.11 Cross layer interactions between the layers

7.5 Summary

In this chapter, space, frequency, and time division multiple access approaches are analysed for supporting simultaneous transmissions in the network. If SDMA approach is used, then spatial re-use supports 5-9 simultaneous routes, however these routes will vary depending on the routes that are active in the network. Spatial re-use is also analysed using three frequency channels. It is shown that three carrier frequencies can support 23 simultaneous routes. However, for guaranteed network access, one frequency per node is required. An alternative approach for multiple simultaneous access is to use TDMA with a separate time slot for each node and each hop. This approach is considered to be the most practical and cost effective for multiple access in a multi-hop ad hoc network.

Nodes in a multi hop route cannot transmit and receive at the same frequency. As a result, throughput per unit time decreases with the number of hops. In order to maintain throughput per unit time, higher order modulation is used to match the number of hops in a route to maximise the throughput per unit time

An enhanced system design is proposed which incorporates TDMA approach and adaptive modulation. The adaptive modulation technique involves cross layer interaction as the route selected at the network layer determines the modulation used at the physical layer.

8. Enhanced System Design Results

8.1 Network performance with throughput maintained

8.1.1 Minimum power routing

A simulation has been carried for the network topology described in figure 6.2 and using scenario A as shown in table 6.3. The MPR scheme is used without maintaining throughput per unit time. Figure 8.1 shows that 45% of the routes are 1-hop routes (directly connected) and uses 6 Mbps; 50% of the routes are 2-hop routes and uses 3 Mbps, and 5% of the routes are 3-hop routes and uses 2 Mbps.

A simulation has been carried out using the same topology with the modulation adapted to maintain the throughput per unit time. Figure 8.2 shows that 88% (including 1-hop, 2-hops and 3-hops) of the routes can maintain maximum throughput per unit time (6 Mbps), 12% of the routes cannot support the maximum throughput because they have insufficient power budget.

From figure 8.1 and figure 8.2, it can be seen that by adapting the modulation to the number of relays in a route can significantly improve the throughput on the majority of routes.

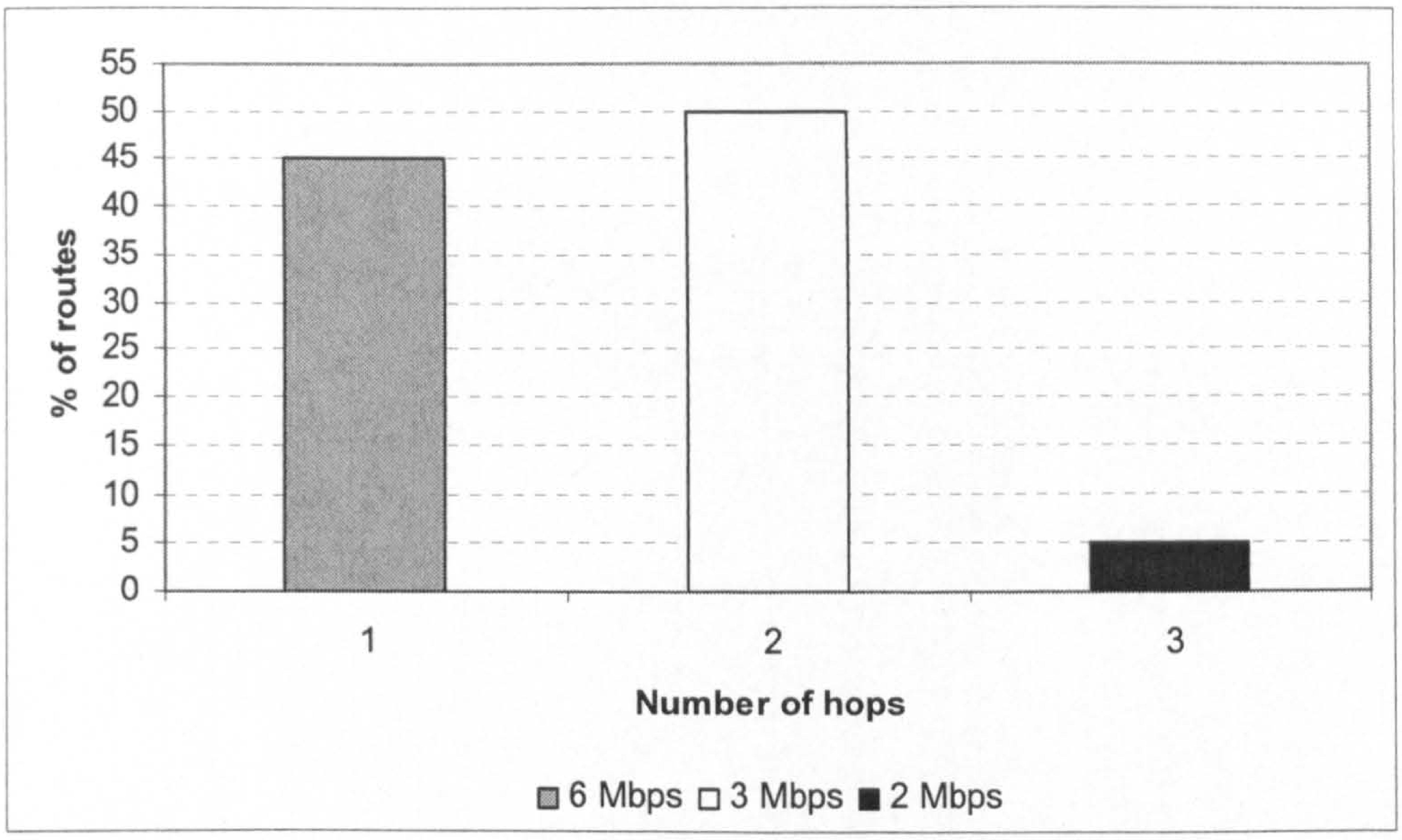


Figure 8.1 Percentage of routes achieving 6 Mbps, 3Mbps and 2Mbps in the MPR scheme

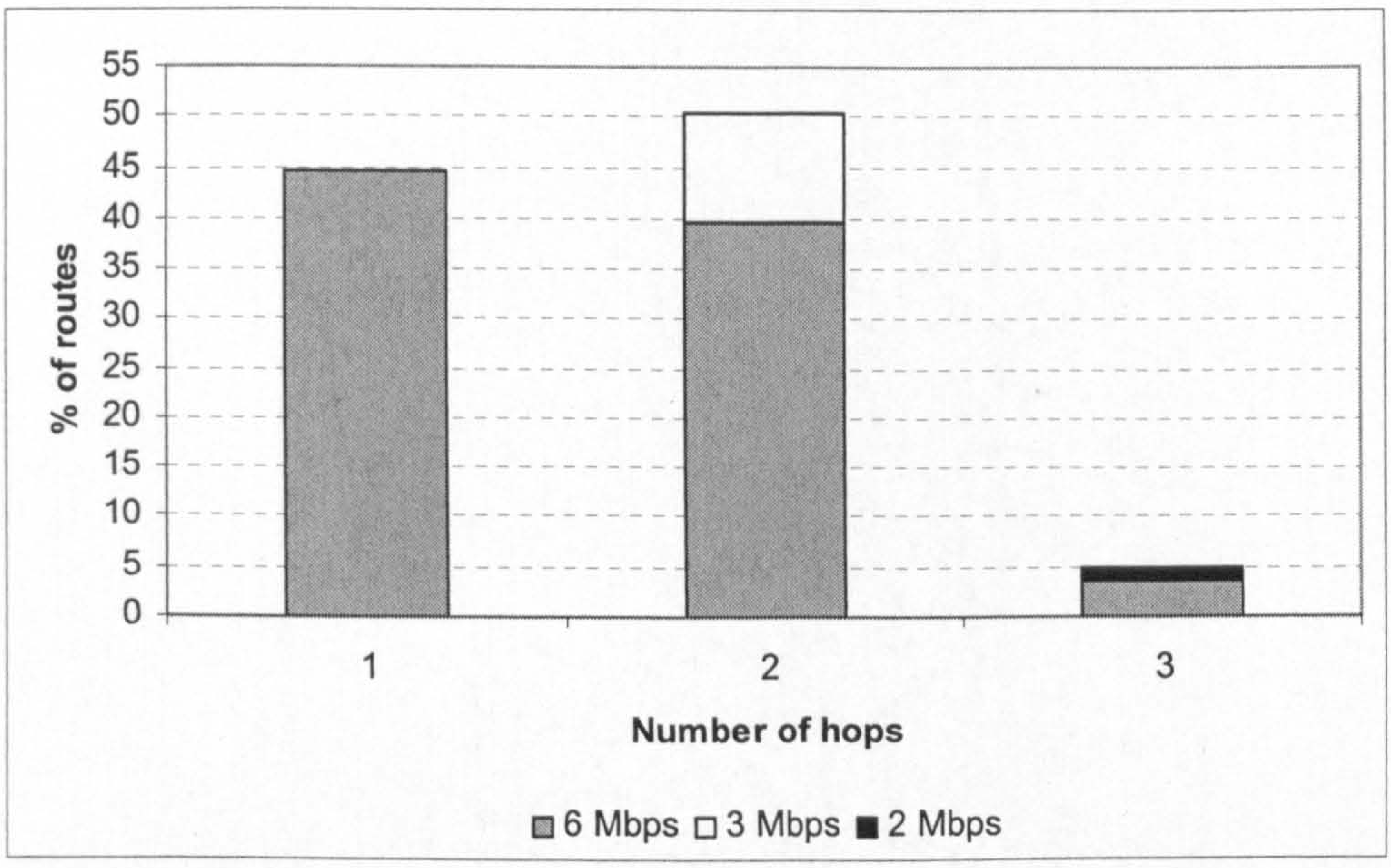


Figure 8.2 Percentage of routes using adaptive modulation in the MPR scheme

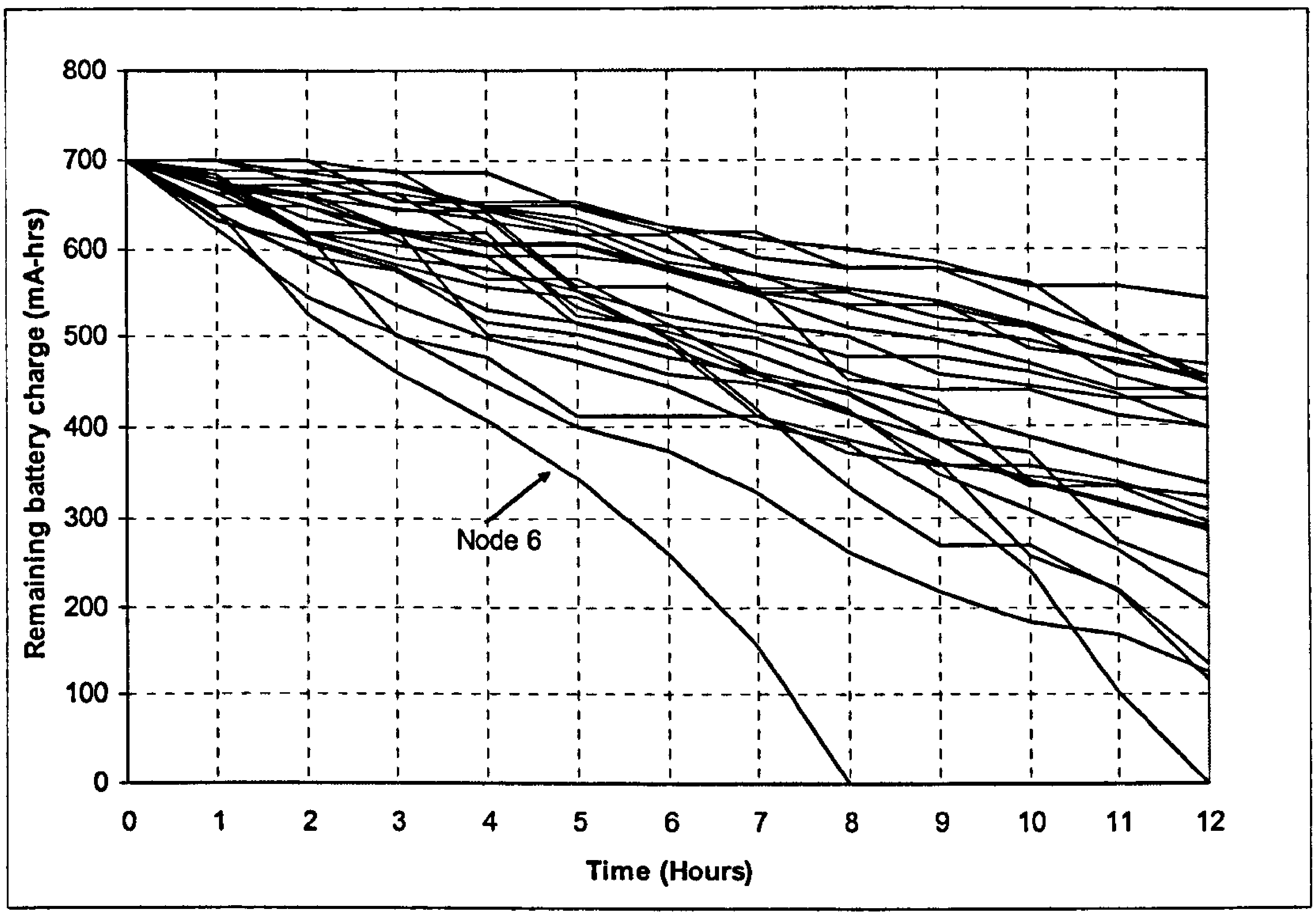


Figure 8.3 Battery charge vs time – MPR scheme using adaptive modulation

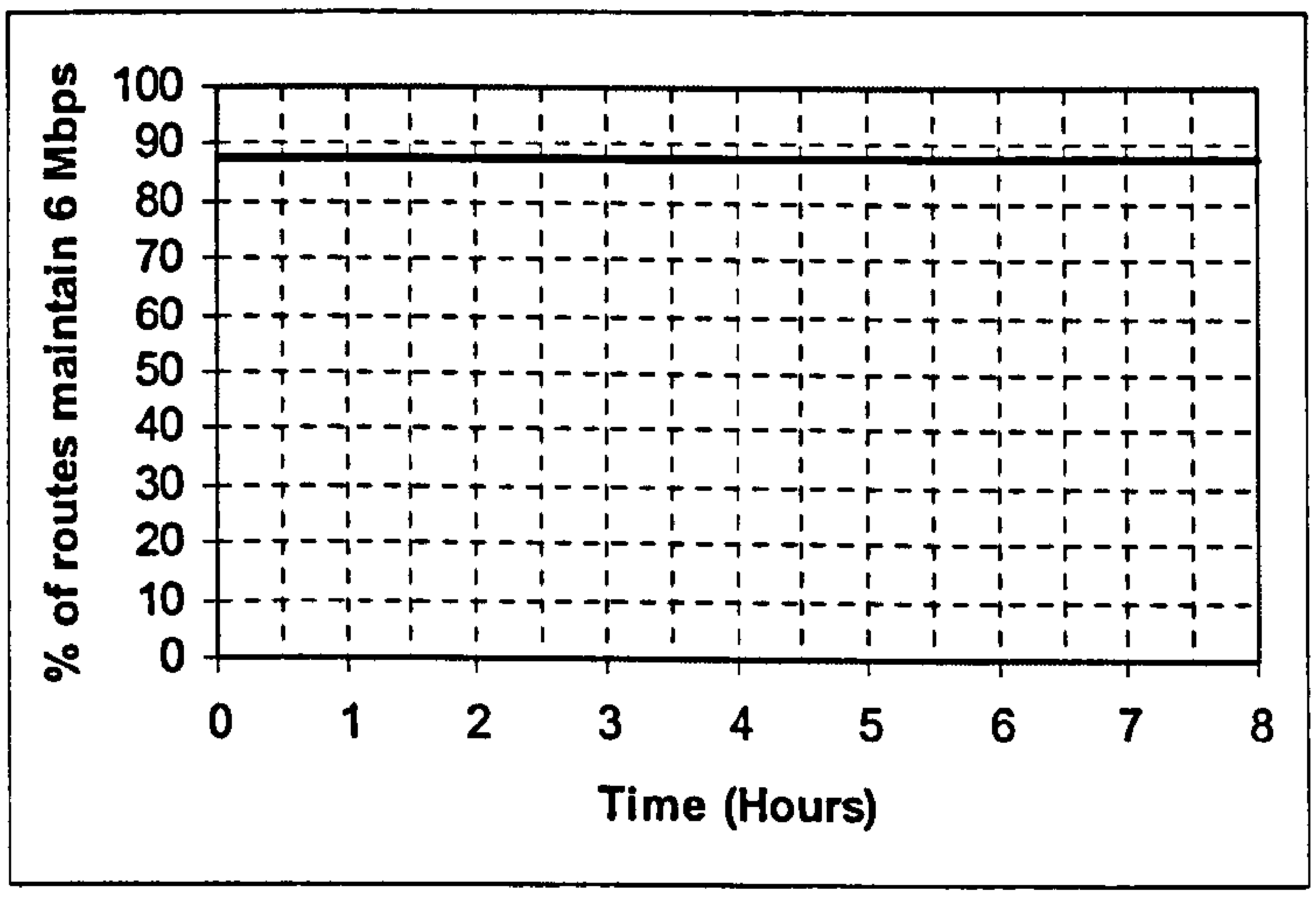


Figure 8.4 Percentage of routes with maximum throughput using adaptive modulation (MPR scheme)

The penalty for this improved throughput is that relay nodes need to use more transmission power and this affects the battery life. Figure 8.3 shows how the residual battery charge varies with time using adaptive modulation. The network lifetime is reduced to 8 hours as compared to the case with MPR with no throughput maintained routes (see figure 6.6). Figure 8.4 shows that 88% of the routes can maintain throughput when all nodes are available.

8.1.2 Battery charge threshold scheme

The network lifetime has also been investigated using throughput maintained with a battery charge threshold of 30%. Figure 8.5 shows that the network lifetime is reduced as compared to 30% threshold with no throughput maintained (see figure 6.10).

Using the same 30% threshold, throughput is measured for each route. Figure 8.6 shows that at the start of the network, 88% of route can support maximum throughput (the same as MPR approach). However, when some of the nodes reach the 30% threshold and are isolated from use as relays, then longer hop routes are selected, and due to insufficient power budget, such routes cannot maintain the throughput. At about 14 hours, all of the isolated nodes are restored to avoid route blockages, and therefore 88% of the routes again able to support the maximum throughput.

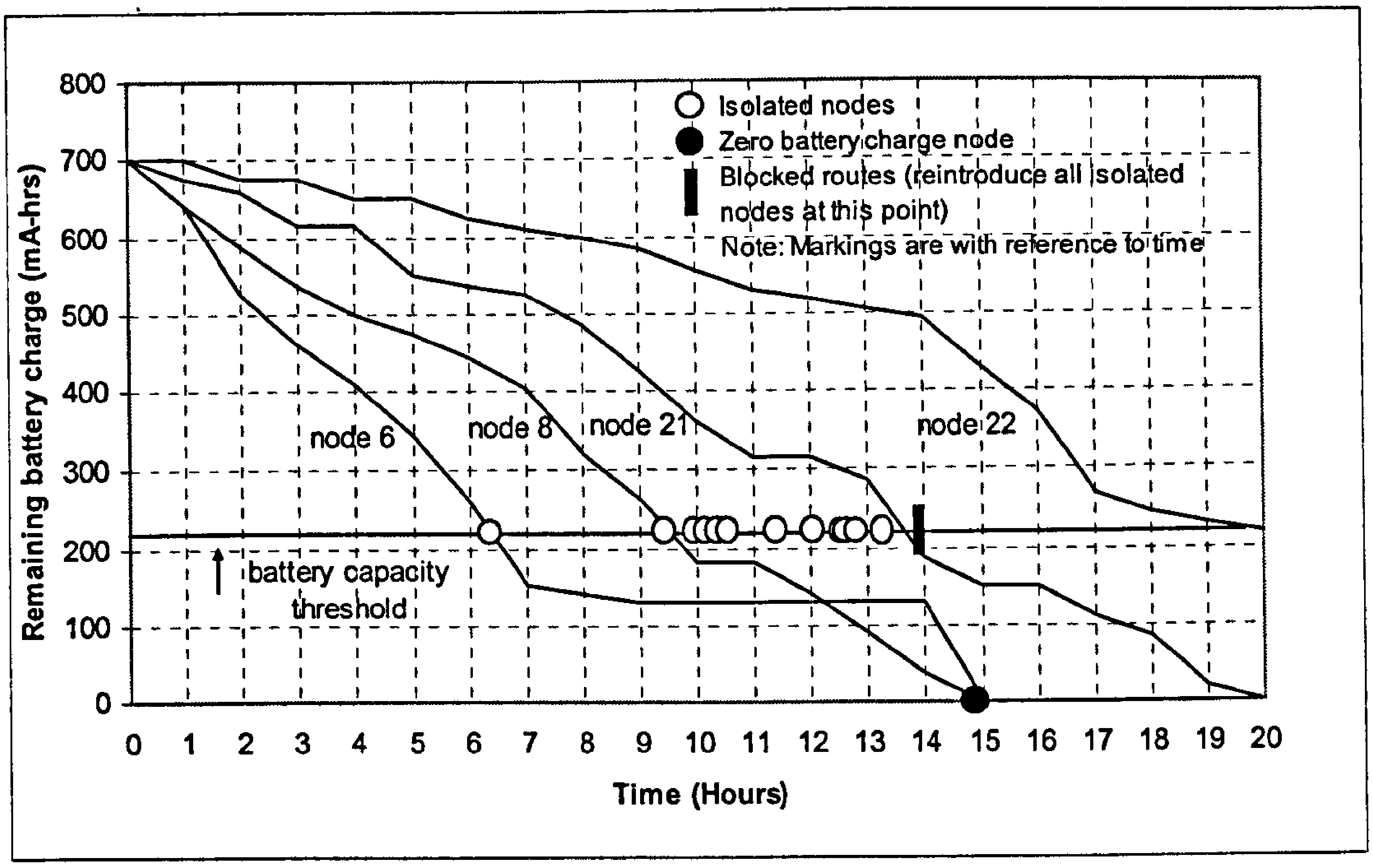


Figure 8.5 Battery charge vs time – throughput enable routes (optimum threshold at 30% of maximum battery charge)

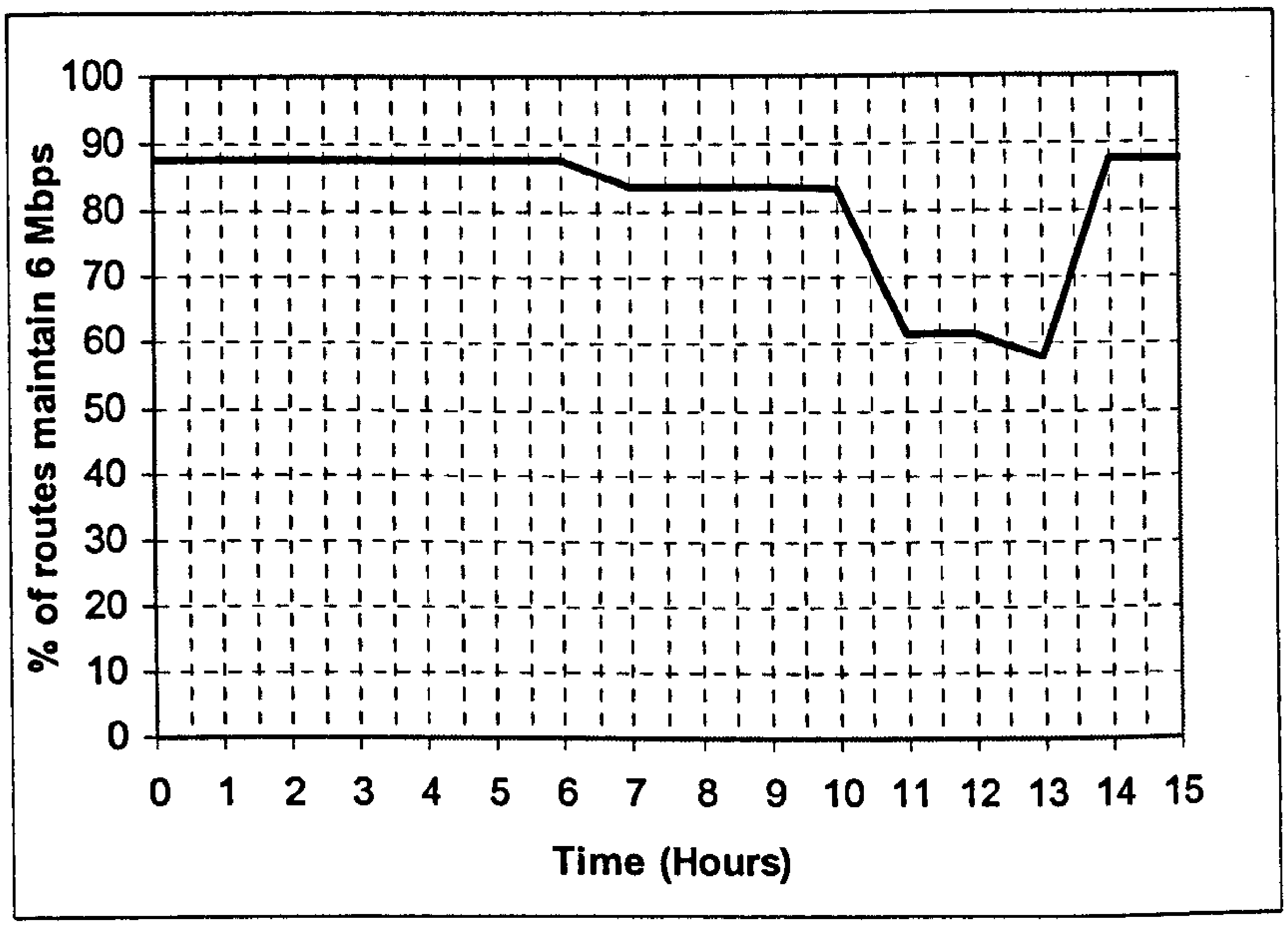


Figure 8.6 Throughput enable routes (optimum threshold at 30% of maximum battery charge)

8.1.3 MPR/MBL approach

The MPR/MBL approach has been evaluated with adaptive modulation to maintain the throughput of the link. Figure 8.7 shows that the network lifetime is reduced from 21 hours (see figure 6.13) to 18 hours.

Figure 8.8 shows the number of throughput enable routes when MPR/MBL approach is used. Similar to the case with MPR and 30% threshold, at the start of the network, 88% of the routes can support maximum throughput. However, the throughput gradually decreases with time. This is due to the fact that most used nodes are not selected as relay nodes and routes are formed with the help of nodes with greater residual battery charge which increases the number of hops in the route.

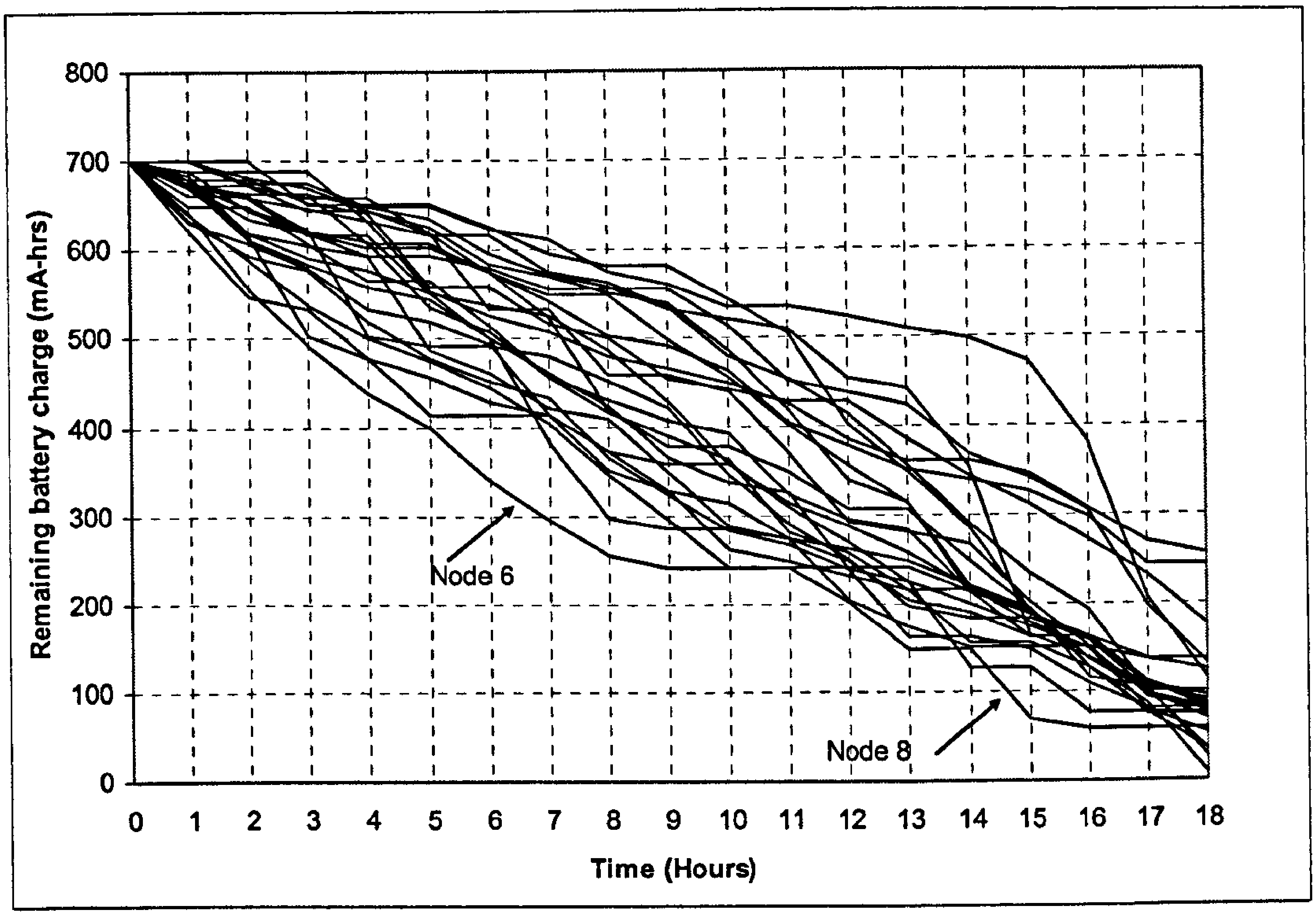


Figure 8.7 Battery charge vs time – throughput enable routes (MPR/MBL approach, cross over point 600mAhrs, a=5)

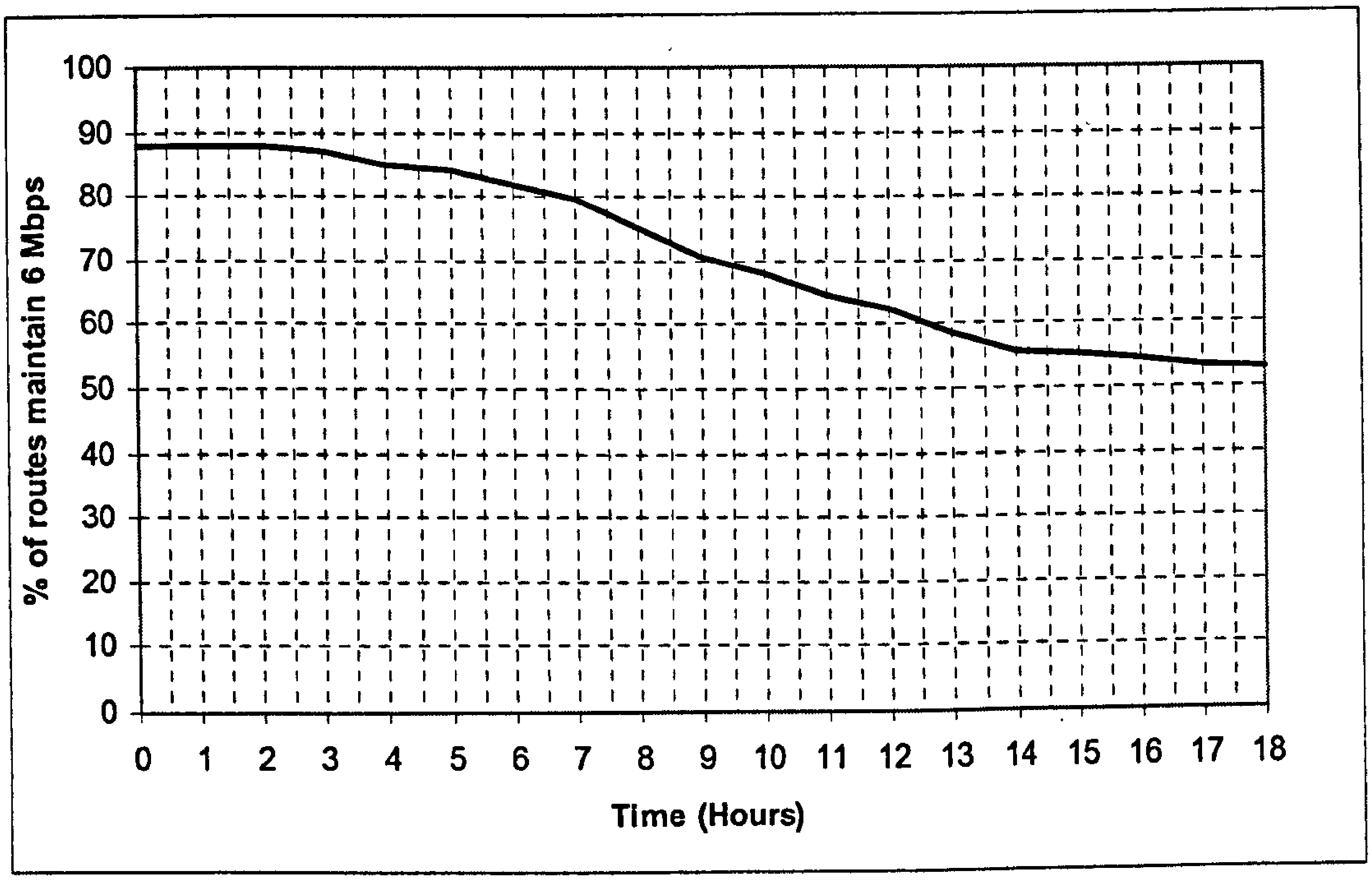


Figure 8.8 Throughput enable routes (MPR/MBL approach)

8.2 Summary

Figure 8.9 shows the comparison of throughput maintained routes between various routing schemes. When throughput per unit time is maintained then the MPR scheme can maintain throughput up to the time when the first node is reached end of battery charge which is 8 hours. The throughput per unit time is maximised up to 15 hours over the case with battery charge threshold scheme. In the MPR/MBL scheme, the throughput is maximised up to 18 hours; however the extra power required for the higher order modulation reduces the network lifetime by 14%.

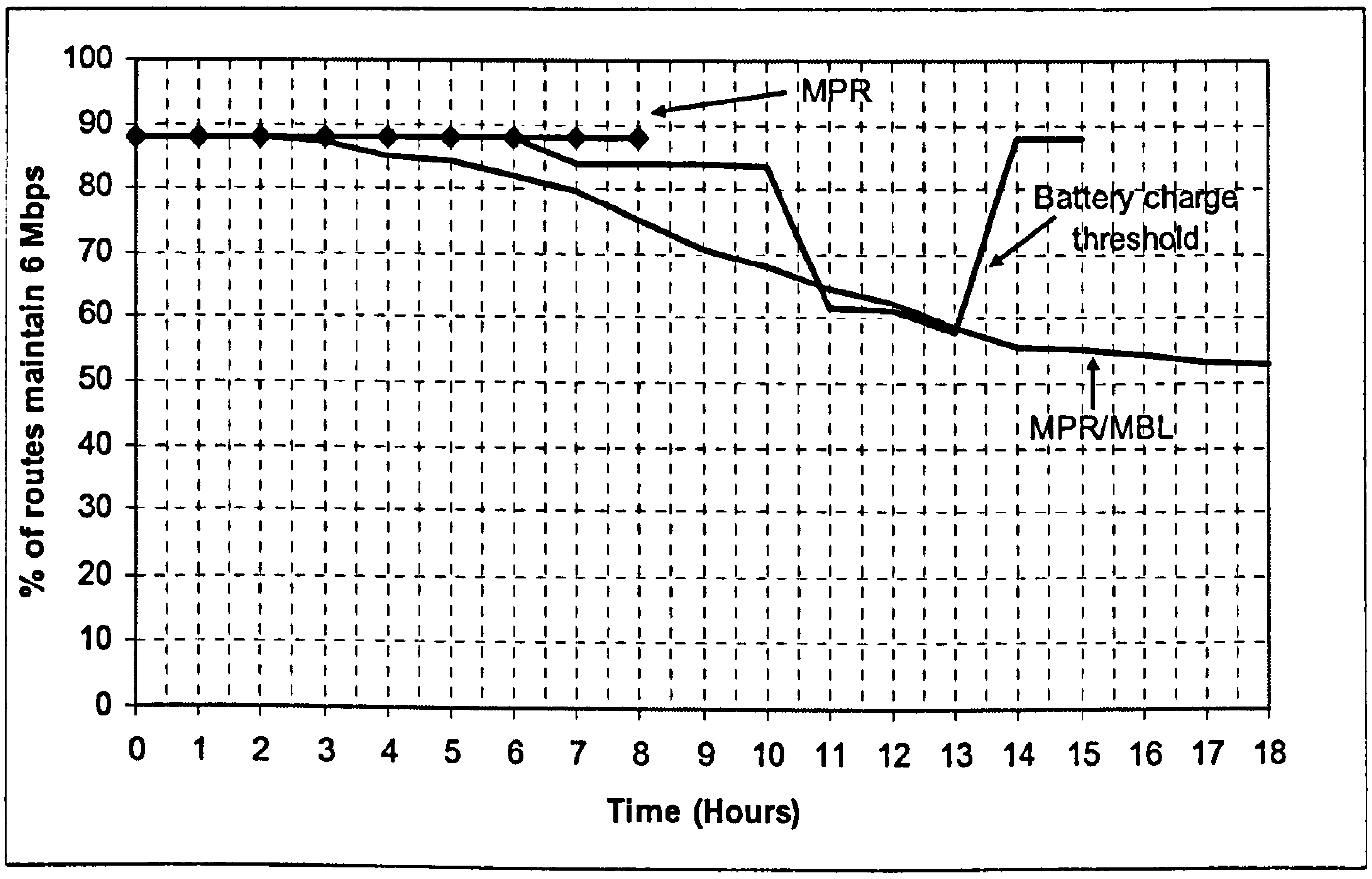


Figure 8.9 Comparison of throughput maintained routes in different routing schemes

9. Conclusion

Ad hoc networks can use relays to form multi hop routes which can circumnavigate high attenuation regions and give a reduction in the power consumption required for data transmission over the end to end route. This thesis has studied multi hop ad hoc network with the aim of identifying the network design that gives the maximum network lifetime. The features required at the each protocol layer in the TCP/IP layered model have been investigated as well as the use of cross layering to improve the performance. A total network solution is described including the transmission and signalling subsystems and a model has been developed to simulate the performance.

At the network layer, a hybrid routing protocol is proposed which can operate in either a proactive mode when nodes are moving rapidly or in reactive mode when nodes are slow or stationary or traffic is transmitted infrequently. The distance vector, source-routed DSR protocol is used in the design. In large networks with mobile nodes, distance vector routing protocols send routing updates to their directly connected neighbours which conserves bandwidth and power. The source routing technique is used in which the end to end routing decision is made at the source node.

Several routing strategies have been examined. The first is based on selecting a route with the minimum end to end power. This minimised the power consumption but may not maximise the network lifetime, as certain nodes may be used more than others if they facilitate a low power path between other nodes, and the battery charge of these nodes can become rapidly depleted. An extension to this protocol has been examined that monitors the battery charge of each node and when the threshold is reached, the node is withdrawn from use as a relay node. An analysis of the threshold has shown that using a threshold value which is 30% of the maximum charge maximises the network lifetime. If route selection is based on the route with the minimum end to end transmission power, then the transmission power needed to transmit between each node in the network needs to be distributed to all nodes so that it can be used by any source node in selecting a route. Similarly the current battery charge status needs to be distributed to all nodes if a battery threshold is introduced into the route selection process. This involves cross layer design as information (transmission power) needs to be passed from the physical layer to the network layer for route selection.

An alternative routing strategy that has been investigated is based on using nodes with the maximum remaining battery charge. This has the advantage that it avoid overusing individual nodes, but it will not always use the lowest power routes so the power consumption may be higher than a minimum power route.

A new hybrid routing scheme has been proposed with the objective of balancing the requirement of minimum power routes and maximum network lifetime. This takes account of both end to end power and the battery charge status on a continual basis throughout the lifetime of the network. When the batteries have a high percentage of their maximum charge, the minimum power route is selected to minimise the power consumption, but as the remaining battery charge for a node decreases, the selection process increasingly weights against the node. Different weighting profiles have been investigated and their performance has been analysed using the network simulation model. This routing scheme requires information about both the transmission power and the battery charge status to be regularly distributed to all nodes in the network.

The transport layer should include ECN and ELN to distinguish segment loss due to congestion and fluctuations in received signal strength. In the absence of these notifications, the standard TCP interprets all segment loss due to congestion, and invokes the congestion avoidance algorithm which decreases the window size and re-transmits lost segment). Such action may not be appropriate if a segment is lost due to high attenuation.

Interference between multiple users can be minimised by using an interference aware medium access scheme. Information about the power being transmitted and received by active nodes is broadcast to all nodes and each node can calculate the maximum power that they can transmit without causing unacceptable interference to existing active nodes. This involves a cross layer interaction between the physical layer and the network layer.

An ad hoc network design has been proposed that supports the features described above. The design includes the data transmission subsystem and also the signalling needed to establish and manage the transmission paths. The solution describes how the signalling can be implemented including the cross layer design features and is based on data from commercially available components. The design uses an out of band beacon channel to implement the signalling system. Each node in the network can access a separate timeslot in the beacon frame and adds its identity, its battery status and the power needed to transmit to neighbour nodes. The beacon frame is continuously transmitted to all nodes in the network and the data rate of this frame is selected to match the speed at which information changes in the network. An information database approach is considered the best approach for managing the cross layer information flows, as it allows data to be transferred between multiple non-adjacent layers.

A 25 node model of the proposed network has been developed and has been used to quantify and compare the performance of the different routing strategies; minimum power, minimum power routing with a battery charge threshold, residual battery charge and the proposed hybrid MPR/MBL scheme. The simulation results show that, assuming batteries with a maximum charge of 700mA-hrs, a network lifetime of 21 hours can be obtained using the proposed MPR/MBL routing scheme. This represents an improvement of 5% over the power aware routing scheme and the residual battery charge scheme, a 31% improvement over the minimum power routing with the optimum battery charge threshold scheme, and an improvement of 133% over the minimum power routing scheme.

In an ad hoc network using a single transmission frequency, the throughput per unit time of the route varies with the number of hops in the route as each node cannot transmit and receive at the same time. It is proposed that adaptive modulation can be used to adjust the throughput in a multi-hop route to maintain the throughput per unit time. This is possible if there is sufficient power budget available to allow the selected modulation. This scheme has been simulated using the simulation model and it is shown that 52% of routes can maintain a throughput per unit time of 6Mbps and the resultant network lifetime is 18 hours.

The options for multiple access have been analysed. Space division multiple access allows a maximum of 5-9 simultaneous routes in the 25 node network that has been modelled. This number depends on the network topology and is limited by interference. If full access for all nodes is required, then frequency division multiple access can be used, but this requires the allocation of one frequency channel per node if full access to be guaranteed. An alternative is to use time division multiple access. This enables full access to be guaranteed but either the data rate per user needs to be reduced or the aggregate transmission rate needs to be increased if the allocated channel has sufficient bandwidth. It is shown how the network can be extended to include adaptive modulation and multiple access.

9.1 Original contributions

The original contributions in this thesis are as follows:

- The optimum remaining battery charge threshold has been calculated for a minimum power routing scheme which stops nodes being used as a relay when their battery energy reaches the threshold.
- A hybrid routing scheme (MPR/MBL) combining both minimum power and maximum battery charge parameters has been proposed. This scheme gives the maximum network lifetime as compared to other reported routing schemes.
- A novel cross-layer scheme is proposed that selects the modulation order to match the number of hops in a route to maximise the throughput per unit time.
- A total network design solution is presented, including both the transmission and signalling subsystem that shows how the novel routing and cross layer features proposed in the thesis can be implemented.

9.2 Future work

Security in the ad hoc network – Beacon frames are broadcast in the network so that information can be shared between the nodes which are essential for link setup. In order to allow certain nodes, authentication is needed so that nodes are first verified before access to the network. Confidentiality and integrity of data is crucial for secure data transmission because data may travel through multiple relays, therefore strong encryption and CRC method is required.

Dynamic time slot allocation – The time slot allocated to the data channel in the current beacon frame is static. In order to efficiently utilize time slots, dynamic time slot allocation can be used. A technique for supporting dynamic time slot allocation needs to be investigated. In addition, further work is needed to investigate dynamic time slot allocation in the beacon channel and a mechanism is required to avoid contention during this process.

Node mobility – The simulation model for 25 node network assumes that the nodes are fixed for the duration of transmission. The network model needs to be extended to incorporate node mobility. Work is needed to determine the impact of higher node mobility on the frequency of the route discovery process, the route stability, the signalling overhead and the lifetime of the network.

Power control – During data transmission, it is assumed that nodes are static and transmission power is fixed. Power control is not implemented during the data transmission. The attenuation information is available; however a mechanism is needed to implement power control.

Handover - When nodes are mobile during transmission, then there is a possibility of link breakage due to the varying path attenuation. As a result, a new route is needed. The RTS/CTS frame exchange to determine a new route increases delay and this may not be acceptable for time sensitive applications. Further work is needed to investigate how the routing protocol and the associated signalling could be adapted to manage a link failure during transmission and the resultant handover.

Scalability – Further work is needed to investigate how the results presented in this thesis scale to a network with a larger number of nodes,

Time dependent simulation – An event driven network model is needed to analyse the data transmission as a function of time and calculate the bit error ratio. Data re-transmission should also be included and this will have an effect on the network lifetime.

References

[1] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, "Advances in packet radio technology", proceedings of the *IEEE*, pp. 1468-1496, November 1978.

[2] N. Abramsom, "Development of the ALOHANET", *IEEE Transactions on Information Theory,* vol. IT-31, no. 2, pp. 119-123, March 1985.

[3] I. M. Jacobs, R. Binder, and E. V. Hoversten, "General purpose packet satellite networks", proceedings of the *IEEE*, pp. 1448-1467, 1978.

[4] B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, et al., "The Past and Future History of the Internet", *Communications of the ACM,* vol. 40, no. 2, pp. 102-108, February 1997.

[5] J. A. Freebersyser and B. Leiner, *A DoD Perspective on Mobile Ad Hoc Networks,* Addison-Wesley Longman, 2001.

[6] W. C. Fifer and F. J. Bruno, "The low-cost packet radio", proceedings of the *IEEE*, pp. 33-42, Jan 1987.

[7] J. M. Rockwell and S. J. Andriole, *Tactical C3 for the Ground Forces,* AFCEA International Press, March 1986.

[8] B. M. Leiner, R. J. Ruther, and A. R. Sastry, "Goals and challenges of the DARPA GloMo program", *IEEE Personal Communications,* vol. 3, no. 6, pp. 34-43, December 1996.

[9] R. B. Adamson, T. Moran, J. Raymond Cole, and M. S. McBeth, "Extended Littoral Battlespace (ELB) Secure Network Voice Gateway", proceedings of the *Military Communications (MILCOM),* New Jersey, USA, pp. 1388-1391, 1999.

[10] "Active IETF Working Groups", available at http://www.ietf.org/html.charters/wg-dir.html, last accessed on 1 May 2008.

[11] "Mobile Ad-hoc Networks (MANET)", available at http://www.ietf.org/html.charters/manet-charter.html, last accessed on 1 May 2008.

[12] "IEEE Std. 802.11-2007, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," IEEE Computer Society.

[13] R. Bruno, M. Conti, and E. Gregori, "Mesh Networks: Commodity Multihop Ad Hoc Networks", *IEEE Communications Magazine,* vol. 43, no. 3, pp. 123-131, March 2005.

[14] "IEEE 802.11s Tutorial: Overview of the Amendment for Wireless Local Area Mesh Networking", available at http://www.ieee802.org/802_tutorials/nov06/802.11s_Tutorial_r5.pdf, last accessed on 15 July 2008.

[15] C.-K. Toh, "Maximum Battery Life Routing to Support Ubiquitous Mobile Computing in Wireless Ad Hoc Networks", *IEEE Communications Magazine,* vol. 39, no. 6, pp. 138-147, June 2001.

[16] A. S. Tanenbaum, *Computer Networks,* Fourth ed., Prentice Hall PTR, 2002.

[17] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-Layer Design for Wireless Networks", *IEEE Communications Magazine,* vol. 41, no. 10, pp. 74-80, October 2003.

[18] M. Conti, G. Maselli, G. Turi, and S. Giordano, "Cross-Layering in Mobile Ad Hoc Network Design", *IEEE Computer Society,* vol. 37, no. 2, pp. 48-51, February 2004.

[19] T. S. Rappaport, A. Annamalai, R. M. Buehrer, and W. H. Tranter, "Wireless Communications: Past Events and a Future Perspective", *IEEE Communications Magazine,* no. 5, Part Anniversary, pp. 148-161, May 2002.

[20] S. Cherry, "Wi-Fi Nodes to Talk Amongst Themselves," in *IEEE Spectrum,* July 2006, pp. 55-56.

[21] C. R. Lin and M. Gerla, "Asynchronous multimedia multihop wireless networks", proceedings of the *IEEE INFOCOM,* Kobe, Japan, pp. 118-125, April 1997.

[22] C. L. Fullmer and J. J. Garcia-Luna-Aceves, "Floor Acquisition Multiple Access (FAMA) for Packet-Radio Networks", proceedings of the *ACM/SIGCOMM,* Cambridge, Massachusetts, USA, pp. 262-273, August 1995.

[23] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A Media Access Protocol for Wireless LAN's", proceedings of the *ACM/SIGCOMM Conference on Communications Architectures, Protocols and Applications,* London, UK, pp. 212-225, August 1994.

[24] I. Chlamtac and A. Farago, "Making transmission schedules immune to topology changes in multi-hop packet radio networks", *IEEE/ACM Transactions on Networking,* vol. 2, no. 1, pp. 23-29, Feburary 1994.

[25] F. Talucci and M. Gerla, "MACA-BI (MACA By Invitation) A Wireless MAC Protocol for High Speed Ad hoc Networking", proceedings of the *IEEE International Conference on Universal Personal Communications (ICUPC)*, San Diego, CA, USA, pp. 913-917, November 1997.

[26] P. Karn, "MACA - A New Channel Access Method for Packet Radio", proceedings of the *ARRL/CRRL Amateur Radio Computer Networking Conference*, pp. 134-140, 1990.

[27] S. Agarwal, R. H. Katz, S. V. Krishnamurthy, and S. K. Dao, "Distributed Power Control in Ad-hoc Wireless Networks", proceedings of the *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, San Diego, USA, pp. F59-F66, September 2001.

[28] J. Gomez, A. T. Campbell, M. Naghshineh, and C. Bisdikian, "Conserving transmission power in wireless ad hoc networks", proceedings of the *IEEE International Conference on Network Protocols (ICNP)*, California, USA, pp. 24-34, November 2001.

[29] J. P. Monks, V. Bharghavan, and W.-m. W. Hwu, "A Power Controlled Multiple Access Protocol for Wireless Packet Networks", proceedings of the *IEEE INFOCOM - The Conference on Computer Communications*, pp. 219-228, April 2001.

[30] M. Krunz, A. Muqattash, and S.-J. Lee, "Transmission Power Control in Wireless Ad Hoc Networks: Challenges, Solutions, and Open Issues", *IEEE Network,* vol. 18, no. 5, pp. 8-14, September/October 2004.

[31] S.-L. Wu, Y.-C. Tseng, and J.-P. Sheu, "Intelligent Medium Access for Mobile Ad Hoc Networks with Busy Tones and Power Control", *IEEE Journal on Selected Areas in Communications,* vol. 18, no. 9, pp. 1647-1657, September 2000.

[32] C.-K. Toh, *Ad Hoc Mobile Wireless Networks: Protocols and Systems*, Prentice Hall PTR, 2002.

[33] E. M. Royer and C. K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless networks", *IEEE Personal Communications,* vol. 6, no. 2, pp. 46-55, April 1999.

[34] C. Perkins, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for mobile computers", proceedings of the *ACM/SIGCOMM Conference on Communications Architectures, Protocols and Applications*, London, UK, pp. 234-44, October 1994.

[35] Z. J. Haas and M. R. Pearlman, "The performance of query control schemes for the zone routing protocol", *IEEE/ACM Transactions on Networking,* vol. 9, no. 4, pp. 427-438, 2001.

[36] D. B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," in *Mobile Computing*, T. Imielinski and H. Korth ed: Kluwer Academic Publishers, 1996, pp. 153-181.

[37] D. Johnson, Y. Hu, and D. Maltz, "RFC 4728: The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4", February 2007.

[38] C.-C. Chiang, H.-K. Wu, W. Liu, and M. Gerla, "Routing in Clustered Multihop, Mobile Wireless Networks with Fading Channel", proceedings of the *IEEE Singapore International Conference on Networks*, Singapore, pp. 197-211, April 1997.

[39] C. Perkins and E. Belding-Royer, "Ad-hoc on-demand distance vector routing", proceedings of the *IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, New Orleans, LA, USA, pp. 90-100, February 1999.

[40] S. Murthy and J. J. Garcia-Luna-Aceves, "An Efficient Routing Protocol for Wireless Networks", *ACM Mobile Networks and Applications: Special issue on Routing in Mobile Communication Networks,* vol. 1, no. 2, pp. 183-97, October 1996.

[41] C.-K. Toh, "Associativity-Based Routing For Ad Hoc Mobile Networks", *Wireless Personal Communications, Special Issue on Mobile Networking and Computing Systems,* vol. 4, no. 2, pp. 103-39, March 1997.

[42] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, et al., "Optimized link state routing protocol for ad hoc networks", proceedings of the *IEEE INMIC*, Lahore, Pakistan, pp. 62-68, December 2001.

[43] T. Clausen and P. Jacquet, "RFC 3626: Optimized Link State Routing Protocol (OLSR)", October 2003.

[44] R. Dube, C. D. Rais, W. Kuang-Yeh, and S. K. Tripathi, "Signal stability-based adaptive routing (SSA) for ad hoc mobile networks", *IEEE Personal Communications,* vol. 4, no. 1, pp. 36-45, February 1997.

[45] G. Pei, M. Gerla, and T.-W. Chen, "Fisheye State Routing: A Routing Scheme for Ad Hoc Wireless Networks", proceedings of the *IEEE International Conference on Communications (ICC)*, New Orleans, LA, USA, pp. 70-74, June 2000.

[46] M. Gerla, X. Hong, and G. Pei, "Landmark Routing for Large Ad Hoc Wireless Networks", proceedings of the *IEEE Globecom*, San Francisco, USA, pp. 1702-6, December 2000.

[47] P. F. Tsuchiya, "The landmark hierarchy: a new hierarchy for routing in very large networks", proceedings of the *Symposium proceedings on communications architectures and protocols*, California, USA, pp. 35-42, August 1988.

[48] "RFC 793: Transmission Control Protocol", September 1981.

[49] W. Stevens, "RFC 2001: TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms", January 2001.

[50] V. Jacobson, "Congestion avoidance and control", proceedings of the *ACM/SIGCOMM*, Stanford, CA, pp. 314-29, August 1988.

[51] S. Kunniyur and R. Srikant, "End-to-End Congestion Control Schemes: Utility Functions, Random Losses and ECN Marks", proceedings of the *IEEE INFOCOM*, Tel Aviv, Israel, pp. 1323-32, March 2000.

[52] H. Balakrishnan and R. H. Katz, "Explicit Loss Notification and Wireless Web Performance", proceedings of the *IEEE Globecom Internet Mini-Conference*, Syndey, Australia, pp. November 1998.

[53] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, and R. H. Katz, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links", *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 756-769, December 1997.

[54] K. Chandran, S. Raghunathan, S. Venkatesan, and R. Prakash, "A feedback-based scheme for improving TCP performance in ad hoc wireless networks", *IEEE Personal Communications*, vol. 8, no. 1, pp. 34-39, February 2001.

[55] J. Liu and S. Singh, "ATCP: TCP for Mobile Ad Hoc Networks", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 7, pp. 1300-1315, July 2001.

[56] K. Xu, S. Bae, S. Lee, and M. Gerla, "TCP behavior across multihop wireless networks and the wired Internet", proceedings of the *ACM International Workshop on Wireless Mobile Multimedia*, Atlanta, Georgia, USA, pp. 41-48, 2002.

[57] S. Xu and T. Saadawi, "Does IEEE 802.11 MAC Protocol Work Well in Multi-hop Wireless Ad Hoc Networks?" *IEEE Communications Magazine,* vol. 39, no. 6, pp. 130-137, June 2001.

[58] S. Xu and T. Saadawi, "Revealing the problems with 802.11 medium access control protocol in multi-hop wireless ad hoc networks", *The International Journal of Computer and Telecommunications Networking,* vol. 38, no. 4, pp. 531-548, March 2002.

[59] M. Allman, V. Paxson, and W. Stevens, "RFC 2581: TCP Congestion Control", April 1999.

[60] R. d. Oliveira and T. Braun, "A Smart TCP Acknowledgement Approach for Multihop Wireless Networks", *IEEE Transactions on Mobile Computing,* vol. 6, no. 2, pp. 192-205, February 2007.

[61] V. Srivastava and M. Motani, "Cross-Layer Design: A Survey and the Road Ahead", *IEEE Communications Magazine,* vol. 43, no. 12, pp. 112-119, December 2005.

[62] V. Kawadia and P. R. Kumar, "Principles and Protocols for Power Control in Wireless Ad Hoc Networks", *IEEE Journal on Selected Areas in Communications,* vol. 23, no. 1, pp. 76-88, January 2005.

[63] S. Floyd, "TCP and Explicit Congestion Notification", *ACM Computer Communication Review,* vol. 24, no. 5, pp. 10-23, October 1994.

[64] K. Ramakrishnan, S. Floyd, and D. Black, "RFC 3168: The Addition of Explicit Congestion Notification (ECN) to IP", September 2001.

[65] "RFC 791: Internet Protocol", September 1981.

[66] K. Nichols, S. Blake, F. Baker, and D. Black, "RFC 2474: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", December 1998.

[67] S. Bradner and V. Paxson, "RFC 2780: IANA Allocation Guidelines For Values In the Internet Protocol and Related Headers", March 2000.

[68] J. d. P. Pavon and S. Choi, "Link adaptation strategy for IEEE 802.11 WLAN via received signal strength measurement", proceedings of the *IEEE International Conference on Communications (ICC),* Anchorage, AK, United States, pp. 1108-1113, May 11-15 2003.

[69] Mohamed-Slim and A. Goldsmith, "Adaptive Modulation over Nakagami Fading Channels", *Wireless Personal Communications,* vol. 13, no. 1-2, pp. 119-143, May 2000.

[70] D. L. Goeckel, "Adaptive Coding for Time-Varrying Channels Using Outdated Fading Estimates", *IEEE Transactions on Communications,* vol. 47, no. 6, pp. 844-855, June 1999.

[71] M. B. Pursley and J. M. Shea, "Adaptive Nonuniform Phase-Shift-Key Modulation for Multimedia Traffic in Wireless Networks", *IEEE Journal on Selected Areas in Communications,* vol. 18, no. 8, pp. 1394-1407, August 2000.

[72] O. Queseth, F. Gessler, and M. Frodigh, "Algorithms for Link Adaptation in GPRS", proceedings of the *IEEE 49th Vehicular Technology Conference,* Houston, TX, USA, pp. 943-947, May 16-20 1999.

[73] W. H. Yuen, H. n. Lee, and T. D. Andersen, "A Simple and Effective Cross Layer Networking System for Mobile Ad Hoc Networks", proceedings of the *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Lisbon, Portugal, pp. 1952-1956, September 2002.

[74] D. Qiao and S. Choi, "Goodput Enhancement of IEEE 802.11a Wireless LAN via Link Adaptation", proceedings of the *IEEE International Conference on Communications (ICC)*, Helsinki, pp. 1995-2000, 2001.

[75] C. Chien, M. B. Srivastava, R. Jain, P. Lettieri, et al., "Adaptive Radio for Multimedia Wireless Links", *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 5, pp. 793-819, May 1999.

[76] M. V. D. Schaar and S. Shankar, "Cross-Layer Wireless: Multimedia Transmission: Challenges, Principles, and New Paradigms", *IEEE Wireless Communications,* vol. 12, no. 4, pp. 50-58, August 2005.

[77] H. Jiang, W. Zhuang, and Z. Shen, "Cross-Layer Design for Resource Allocation in 3G Wirless Networks and Beyond", *IEEE Communications Magazine,* vol. 43, no. 12, pp. 120-126, December 2005.

[78] P. Bender, P. Black, M. Grob, R. Pandovani, et al., "CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users", *IEEE Communications Magazine,* vol. 38, no. 7, pp. 70-77, July 2000.

[79] T. S. Rappaport, *Wireless Communications: Principles and Practice,* Prentice Hall, 2002.

[80] A. Muqattash and M. Krunz, "Power Controlled Dual Channel (PCDC) Medium Access Protocol for Wireless Ad Hoc Networks", proceedings of the *IEEE INFOCOM*, San Francisco, California, pp. 470-480, March 2003.

[81] V. Bhuvaneshwar, M. Krunz, and A. Muqattash, "CONSET: A Cross-Layer Power Aware Protocol for Mobile Ad Hoc Networks", proceedings of the *IEEE International Conference on Communications (ICC)*, Paris, France, pp. 4067-4071, 2004.

[82] L. Qin and T. Kunz, "Increasing Packet Delivery Ratio in DSR by Link Prediction", proceedings of the *36th International Conference on System Sciences (HICSS'03)*, Hawaii, pp. 300.1, 2003.

[83] P. P. Pham, S. Perreau, and A. Jayasuriya, "New Cross-Layer Design Approach to Ad Hoc Networks Using Rayleigh Fading", *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 28-39, January 2005.

[84] G. Wu, Y. Bai, J. Lai, and A. Ogielski, "Interactions between TCP and RLP in Wireless Internet", proceedings of the *IEEE Globecom*, Rio de Janeiro, Brazil, pp. 661-666, December 1999.

[85] V. T. Raisinghani and S. Iyer, "Cross-Layer Feedback Architecture for Mobile Device Protocol Stacks", *IEEE Communications Magazine*, vol. 44, no. 1, pp. 85-92, January 2006.

[86] P. Sudame and B. R. Badrinath, "On Providing Support for Protocol Adaptation in Mobile Wireless Networks", *Mobile Networks and Applications*, vol. 6, no. 1, pp. 43-55, 2001.

[87] Q. Wang and M. A. Abu-Rgheff, "Cross-Layer Signalling for Next-Generation Wireless Systems", proceedings of the *IEEE Wireless Communications & Networking Conference*, New Orleans, Louisiana, USA, pp. 1084-89, March 2003.

[88] B.-J. Kim, "A network service providing wireless channel information for adaptive mobile applications: part I: proposal", proceedings of the *IEEE International Conference on Communications (ICC)*, Helsinki, Finland, pp. 1345-51, June 2001.

[89] K. Chen, S. H. Shah, and K. Nahrstedt, "Cross-Layer Design for Data Accessibility in Mobile Ad Hoc Networks", *Wireless Personal Communications*, vol. 21, no. 1, pp. 49-76, 2002.

[90] R. Winter, J. H. Schiller, N. Nikaein, and C. Bonnet, "CrossTalk: Cross-Layer Decision Support Based on Global Knowledge", *IEEE Communications Magazine,* vol. 44, no. 1, pp. 99-99, 2006.

[91] "RFC 792: Internet Control Message Protocol", September 1981.

[92] V. Rodoplu and T. H. Meng, "Minimum Energy Mobile Wireless Networks", *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 8, pp. 1333-1344, August 1999.

[93] D. Khan, P. Ball, and G. Childs, "A Holistic Cross Layer Design Approach for Multi-Hop Mobile Ad Hoc Networks", proceedings of the *Fifth International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP),* Patras, Greece, pp. 220-223, 19-21 July, 2006.

[94] M. Schwartz, *Mobile Wireless Communications,* Cambridge University Press, 2004.

[95] D. Khan, P. Ball, and G. Childs, "A Holistic Cross Layer Design Approach for Multi-Hop Mobile Ad Hoc Networks", proceedings of the *Fifth International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP),* Patras, Greece, pp. 220-223, July 2006.

[96] A. Michail and A. Ephremides, "Energy Efficient Routing for Connection-Oriented Traffic in Ad-Hoc Wireless Networks", proceedings of the *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications,* London, UK, pp. 762-766, September 2000.

[97] M. Maleki, K. Dantu, and M. Pedram, "Power-aware Source Routing Protocol for Mobile Ad Hoc Networks", proceedings of the *International Symposium on Low Power Electronics and Design (ISLPED),* California, USA, pp. 72-75, August 2002.

[98] V. Marbukh and M. W. Subbarao, "Framework for Maximum Survivability Routing for a MANET", proceedings of the *21st Century Military Communications (MILCOM),* Los Angeles, USA, pp. 282-286, October 2000.

[99] "Indium Gallium Phosphorus HBT (WLAN Power Amplifier) - MMG2401NR2," Freescale Semiconductor, Technical Data May 2006.

[100] "IEEE Std. 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," IEEE Computer Society 1999.

[101] "ETS 300 328: Technical characteristics and test conditions for data transmission equipment operating in the 2.4GHz ISM band and using spread

spectrum modulation techniques," European Telecommunication Standards Institute November 1996.

[102] Hans-Otto, "Beyond the Law," in *IET Communications Engineer*, April/May 2007, pp. 37-39.

[103] "Conexant Systems - Single-Chip WLAN Radio - CX53111", 2005, available at http://www.conexant.com/servlets/DownloadServlet/PBR-200883-001.pdf?docid=884&revid=1, last accessed on 23 April 2008.

[104] "3Com OfficeConnect Wireless 54 Mbps 11g Access Point, WL-524", 2006, available at http://www.3com.com/other/pdfs/products/en_US/3com_400988.pdf, last accessed on 22 April 2008.

[105] S. Singh, M. Woo, and C. S. Raghavendra, "Power-aware routing in mobile ad hoc networks", proceedings of the *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, Dallas, Texas, USA, pp. 181-190, 1998.

[106] W. Stallings, *Data and Computer Communications*, 5th ed., Prentice-Hall, Inc., 1997.

[107] S. Haykin and M. Moher, *Modern Wireless Communications*, International ed., Prentice Hall, 2004.

[108] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, 3rd ed., John Wiley and Sons Ltd., 2002.

[109] "Cisco Aironet 802.11A/B/G Wireless CardBus Adapter", available at http://www.cisco.com/en/US/prod/collateral/wireless/ps6442/ps4555/ps5818/product_data_sheet09186a00801ebc29.pdf, last accessed on 22 April 2008.

[110] "NXP Semiconductors - BGW211 Low-power WLAN SiP", 2007, available at http://www.activecomponents.com/pdf/datasheets/BGW211.pdf, last accessed on 23 April 2008.

[111] "ST - Single chip 802.11b/g WLAN radio (STLC4560)", available at http://www.st.com/stonline/products/literature/bd/14302.pdf, last accessed on 23 April 2008.

[112] "IEEE Std. 802.11g, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band," IEEE Computer Society 2003.

[113] A. Nasipuri, J. Zhuang, and S. R. Das, "A Multichannel CSMA MAC Protocol for Multihop Wireless Networks", proceedings of the *IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, Louisiana, pp. 1402-1406, September 1999.

[114] N. Jain, S. R. Das, and A. Nasipuri, "A Multichannel CSMA MAC Protocol with Receiver-Based Channel Selection for Multihop Wireless Networks", proceedings of the *International Conference on Computer Communications and Networks (ICCCN)*, Scottsdale, Arizona USA, pp. 432-439, October 2001.

[115] S.-L. Wu, Y.-C. Tseng, C.-Y. Lin, and J.-P. Sheu, "A Multi-channel MAC Protocol with Power Control for Multi-hop Mobile Ad Hoc Networks", *The Computer Journal*, vol. 45, no. 1, pp. 101-110, 2002.

[116] J. So and N. Vaidya, "Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals Using a Single Transceiver", proceedings of the *ACM/SIGMOBILE MobiHoc*, Roppongi, Japan, pp. 222-233, May 2004.

[117] "IEEE Std 802.16j: Publication History", available at http://www.ieee802.org/16/pubs/80216j.html, last accessed on 10 May 2008.

[118] "Out-of-band relay clarification", available at http://wirelessman.org/relay/, last accessed on 10 May 2008.

[119] "In-band and out-of-band definition clean up", available at http://wirelessman.org/relay/, last accessed on 10 May 2008.

[120] "Aruba Networks: WLAN RF Architecture Primer", available at http://www.arubanetworks.com/pdf/technology/whitepapers/wp_RFARCH.pdf, last accessed on 17 June 2008.

[121] Xirrus, "Deploying High Performance Wi-Fi Networks", available at http://www.xirrus.com/library/, last accessed on 14 May 2008.

[122] J. G. Proakis and M. Salehi, *Communication Systems Engineering*, 2nd ed., Prentice Hall, April 2003.

Published work

1. D. Khan, P. Ball, and G. Childs, "Extending the Lifetime of Multi Hop Ad hoc Networks by Managing the Use of Relay Nodes", proceedings of the Sixth International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Graz, Austria, pp. 657-661, July 2008.

2. D. Khan, P. Ball, and G. Childs, "A Cross Layer Design Approach for Multi-hop Mobile Ad Hoc Networks", The Mediterranean Journal of Computers and Networks (MEDJCN), vol. 3, no. 3, pp. 91-99, July 2007.

3. D. Khan, P. Ball, and G. Childs, "A Holistic Cross Layer Design Approach for Multi-Hop Mobile Ad Hoc Networks", proceedings of the Fifth International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Patras, Greece, pp. 220-223, 19-21 July 2006.

Appendices

A Neighbour discovery process

Neighbour discovery is the first process of routing in the ad hoc network which defines how nodes are connected in the network and what information is available to share with other nodes. In the neighbour discovery process, each node gathers information about other nodes in the network, and with the knowledge of neighbouring nodes, routing is performed.

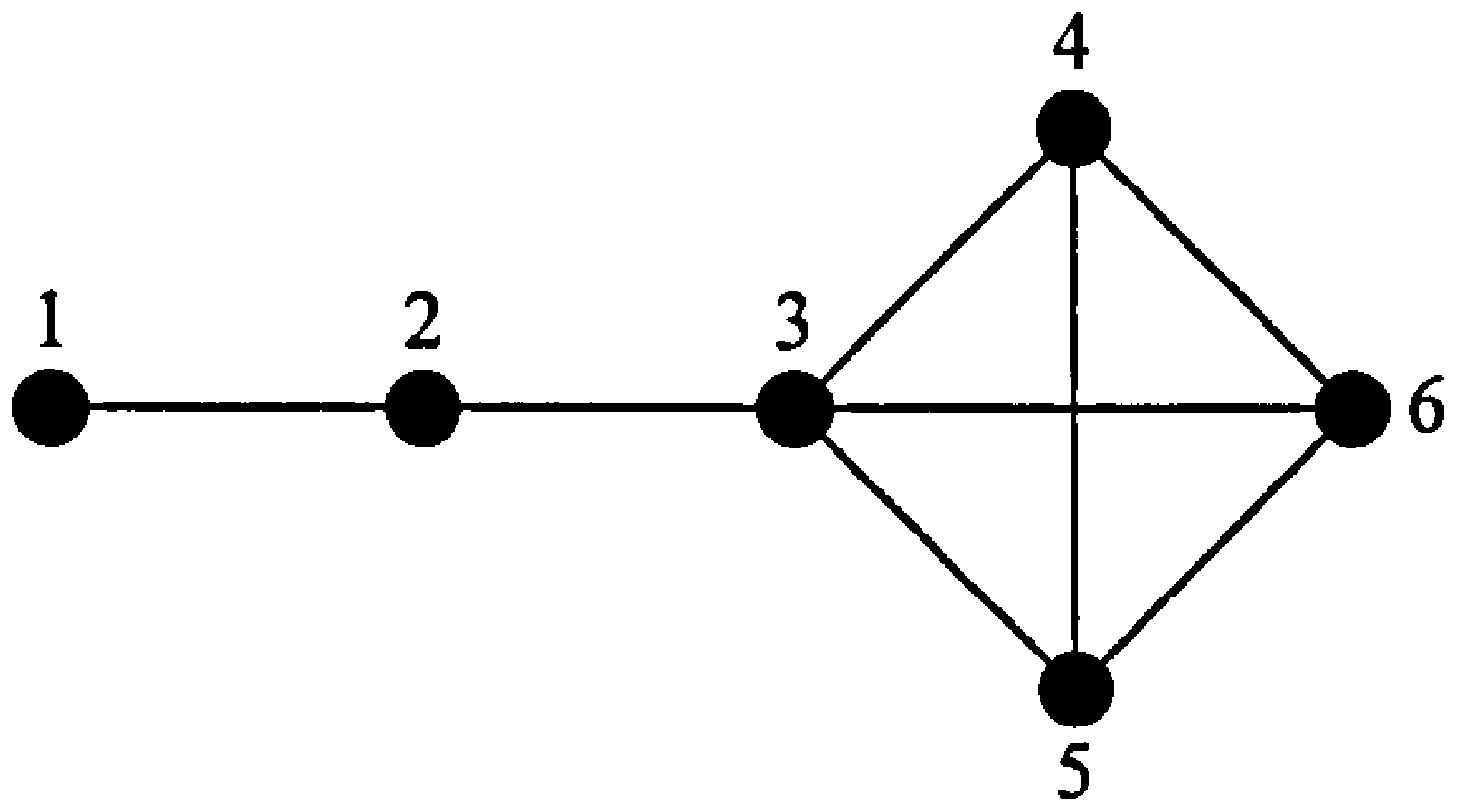


Figure A.1 Example network for neighbour discovery

Figure A.1 shows the example network which is used to describe the process of neighbour discovery. In this network, it is assumed that all the nodes are freshly started and nodes 3, 4, 5 and 6 form a mesh network; however nodes 1 and 2 are connected in line with node 3. The neighbour discovery process consists of three steps:

i) each node broadcasts its own information,
ii) each node broadcasts information about directly connected neighbour nodes,
iii) each node forwards neighbour node beacon frames to other neighbours when they are received.

In the first step, if a node wants to communicate but does not hear any on-going beacon transmission, then it sets the timing and beacon bit rate, then adds the following information to the time slot: a node ID, its battery status, the resource discovery mode it wants to use, and received power and then broadcasts this information at the fixed beacon power. Other nodes when join the network and perform initial resource discovery will hear this broadcast and then synchronise their timing and bit rate.

(i) Broadcast own node information

At the start of the neighbour discovery process, nodes in the network are not aware of other neighbours in the network. Each node broadcasts its own information in the network. The information contains node ID, minimum power, available battery and received power. After this process, all the nodes are aware of their directly connected neighbours. For example, node 1 is aware of node 2, and node 3 is aware of nodes 2, 4, 5 and 6 (all directly connected nodes), see Figure A.2.

NID=1

NN	MP	RP	B
2	1-2	x	x

NID=2

NN	MP	RP	B
1	2-1	x	x
3	2-3	x	x

NID=3

NN	MP	RP	B
2	3-2	x	x
4	3-4	x	x
5	3-5	x	x
6	3-6	x	x

NID=4

NN	MP	RP	B
3	4-3	x	x
5	4-5	x	x
6	4-6	x	x

NID=5

NN	MP	RP	B
3	5-3	x	x
4	5-4	x	x
6	5-6	x	x

NID=6

NN	MP	RP	B
3	6-3	x	x
4	6-4	x	x
5	6-5	x	x

NID	Node ID
NN	Neighbour node
MP	Minimum power required between the nodes
RP	Received power
B	Available battery of the neighbour node

Figure A.2 Neighbour table after the first broadcast

(ii) Broadcast directly connected neighbour node's information

In the second broadcast, each node broadcasts directly connected neighbour node information in the beacon. For example, node 2 broadcasts information about directly connected nodes (nodes 1 and 3), and similarly node 2 receives broadcasts from its directly connected neighbours. In this way, node 1 is able to know the neighbours of its directly connected node, hence node 1 is aware of node 3.

(iii) Forward neighbour node beacon frames to other neighbours

Nodes in a wireless network may be hidden possibly due to obstacles and hence broadcasts may be confined to one region. Broadcasts can be propagated in the network when each node forwards beacon frames of the other neighbouring nodes. In the example, node 1 cannot detect broadcasts from nodes 3, 4, 5 and 6. However, if node 2 forwards received beacon frames (such as node 3 beacon frame), then node 1 is able to learn about nodes 3-6. Hence after a few iteration, nodes are able to successfully distribute information to all other nodes and know the complete topology of the network.

B Power budget for the data channel

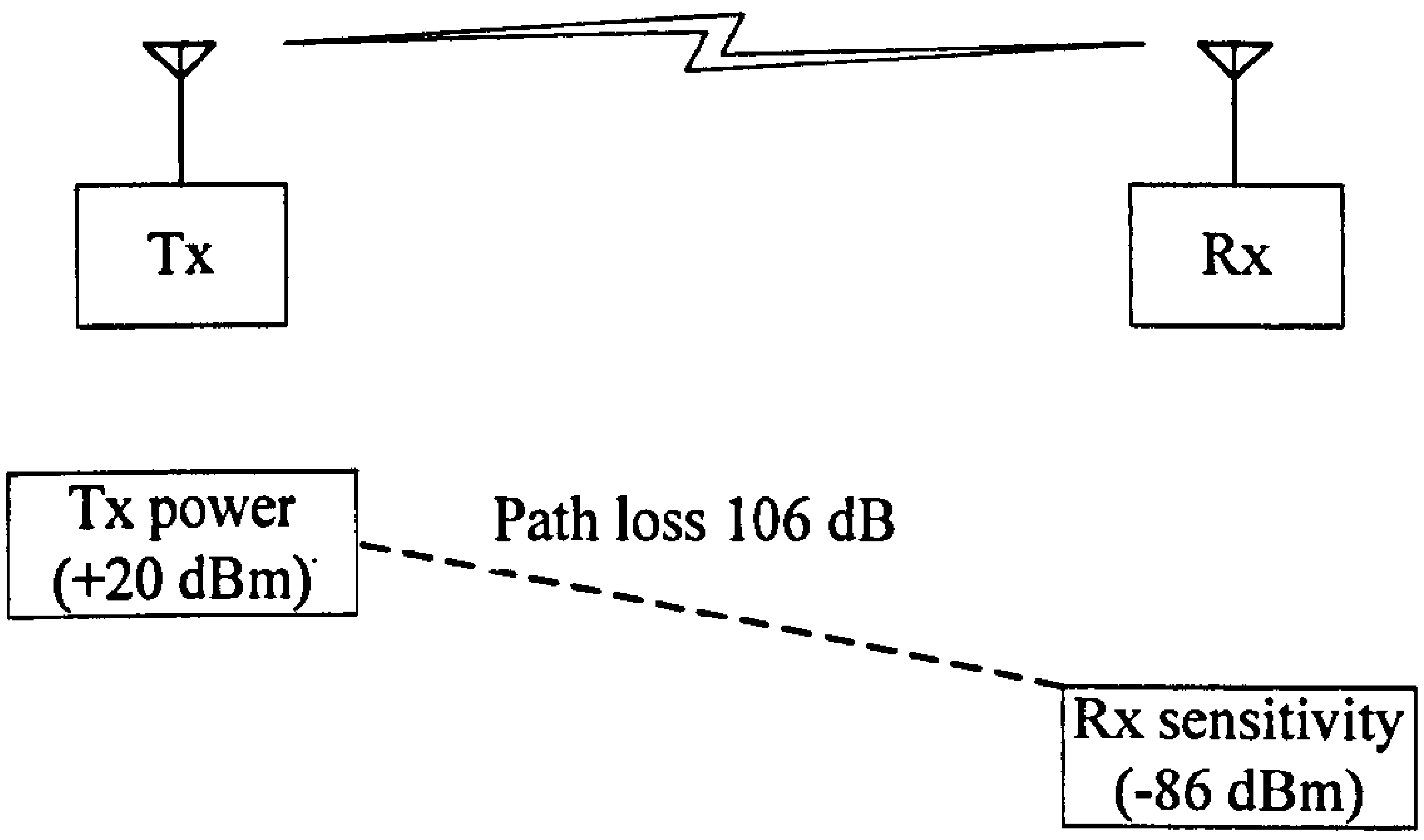


Figure B.1 Power budget for the data channel

The maximum allowed power is set to +20 dBm based on the value in the IEEE 802.11 specifications [12]. An average receiver sensitivity value of -86 dBm is based on commercially available datasheets [104, 109] for equipment supporting the IEEE 802.11 standard. The receiver and transmitter parameters give a maximum path loss of 106 dB.

C Power budget for the beacon channel

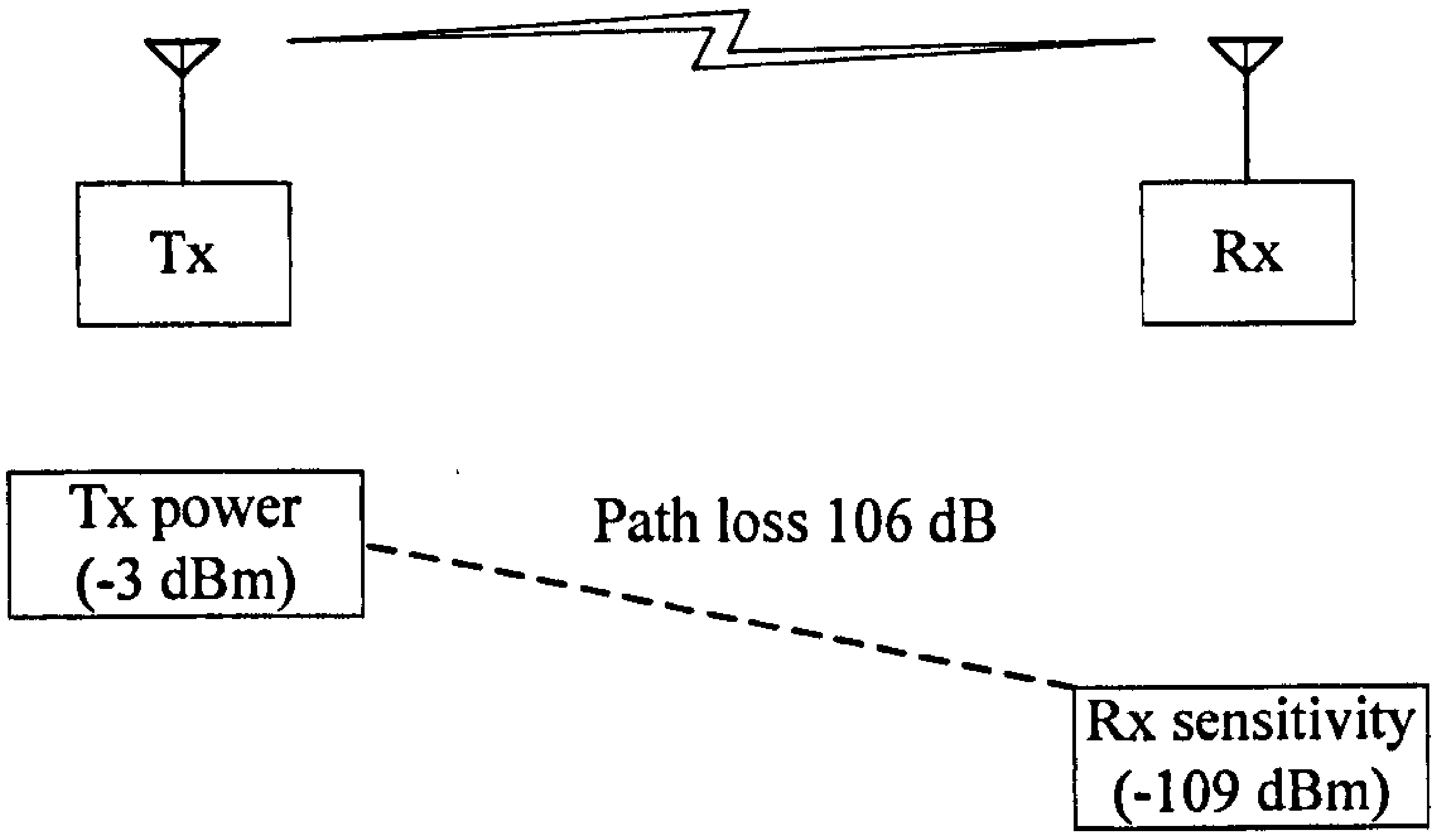


Figure C.1 Power budget for the beacon channel

Using figure 5.2, the path loss of 106 dB covers a distance of 20m using the Ground Reflection Model with 50dB clutter factor and 11dB fade margin. With a distance of 20m for each beacon transmission, minimum three relays are required to cover the end to end distance of 70m in the network size of 50m x 50m.

The beacon bit rate is 32kbps. The sensitivity of the receiver at 32kbps beacon bit rate is 23dB more sensitive than the receiver sensitivity at 6Mbps (-86dBm). This is calculated by 10log (6Mbps/32kbps).

Using the path loss of +106dB and receiver sensitivity of -109dBm, the transmission power is -3dBm.

D Beacon bit rate calculations

Figure D.1 shows the beacon frame format.

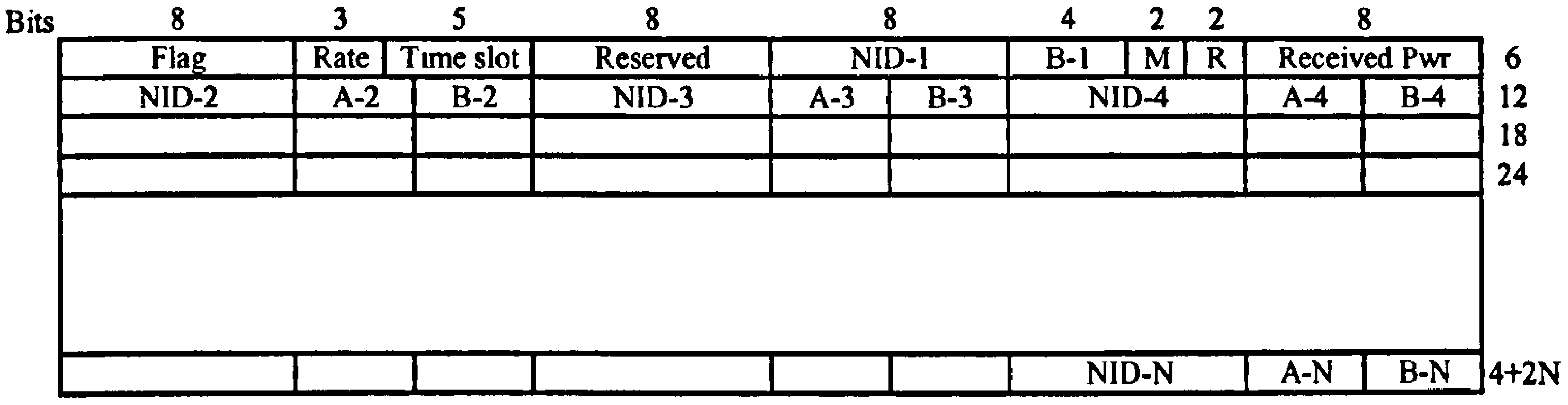


Figure D.1 Beacon frame format

Beacon frame size = 8(4+2N) bits

Total beacon transmissions for N nodes = 8N(4+2N) bits

Rate of transmission = velocity of node / maximum distance allowed

= 0.27V/D, where V in km/hr and D in meters

where 0.27 is a conversion factor from m/s to km/hr

Bit rate (bps) = 8N(4+2N) x 0.27 V/D

= 2.22N(4+2N) V/D

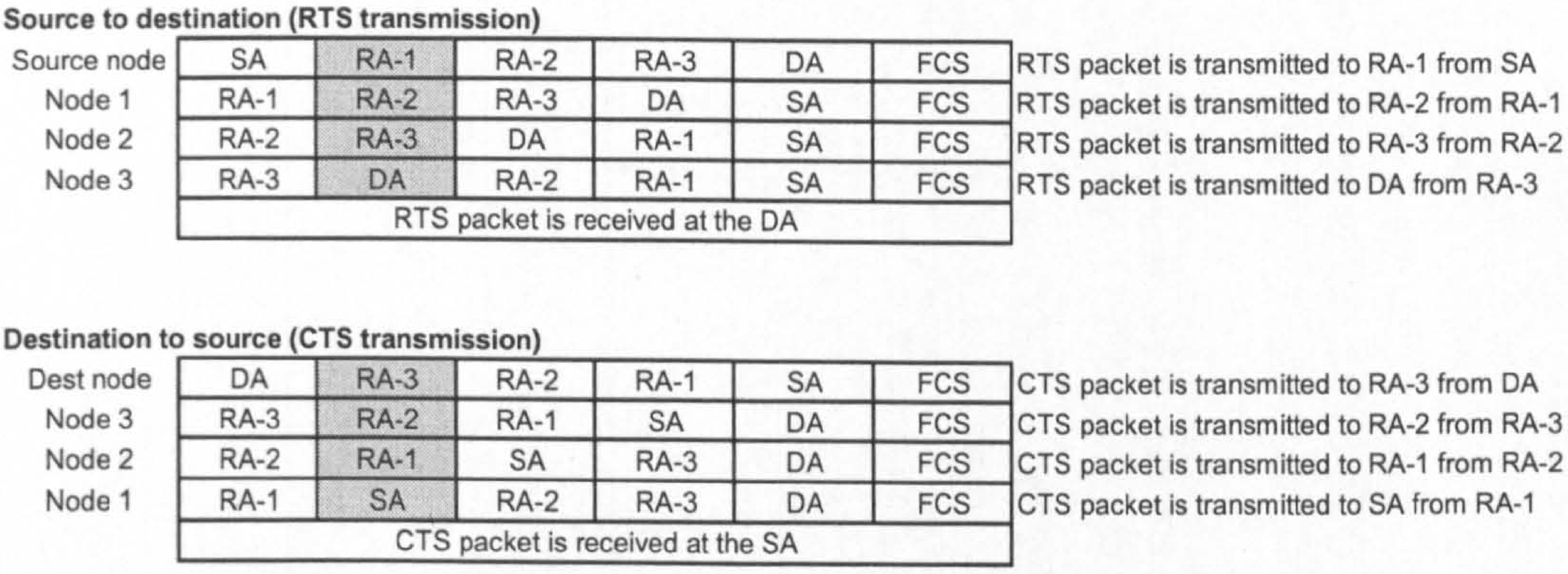
Three beacon transmissions are needed to ensure end to end transmission over the maximum distance between the nodes. Hence,

Bit rate (bps) = 3[2.22N(4+2N) V/D]

= 6.66N(4+2N) V/D

E RTS/CTS process

Figure E.1 shows the RTS/CTS transmissions and Figure E.2 shows the process of RTS/CTS in an ad hoc network. When RTS is transmitted from the source node at the selected transmission power, there is a possibility that other nodes may receive a RTS frame. However by matching the second address field, only the designated node (either the relay or the destination node) can respond to the RTS frame. The RTS frame also signals to isolated nodes to become active when an alternative route is not available and relay data from the requested source node.



Source to destination (RTS transmission)

Source node	SA	RA-1	RA-2	RA-3	DA	FCS	RTS packet is transmitted to RA-1 from SA
Node 1	RA-1	RA-2	RA-3	DA	SA	FCS	RTS packet is transmitted to RA-2 from RA-1
Node 2	RA-2	RA-3	DA	RA-1	SA	FCS	RTS packet is transmitted to RA-3 from RA-2
Node 3	RA-3	DA	RA-2	RA-1	SA	FCS	RTS packet is transmitted to DA from RA-3
	RTS packet is received at the DA						

Destination to source (CTS transmission)

Dest node	DA	RA-3	RA-2	RA-1	SA	FCS	CTS packet is transmitted to RA-3 from DA
Node 3	RA-3	RA-2	RA-1	SA	DA	FCS	CTS packet is transmitted to RA-2 from RA-3
Node 2	RA-2	RA-1	SA	RA-3	DA	FCS	CTS packet is transmitted to RA-1 from RA-2
Node 1	RA-1	SA	RA-2	RA-3	DA	FCS	CTS packet is transmitted to SA from RA-1
	CTS packet is received at the SA						

Figure E.1 RTS/CTS transmissions

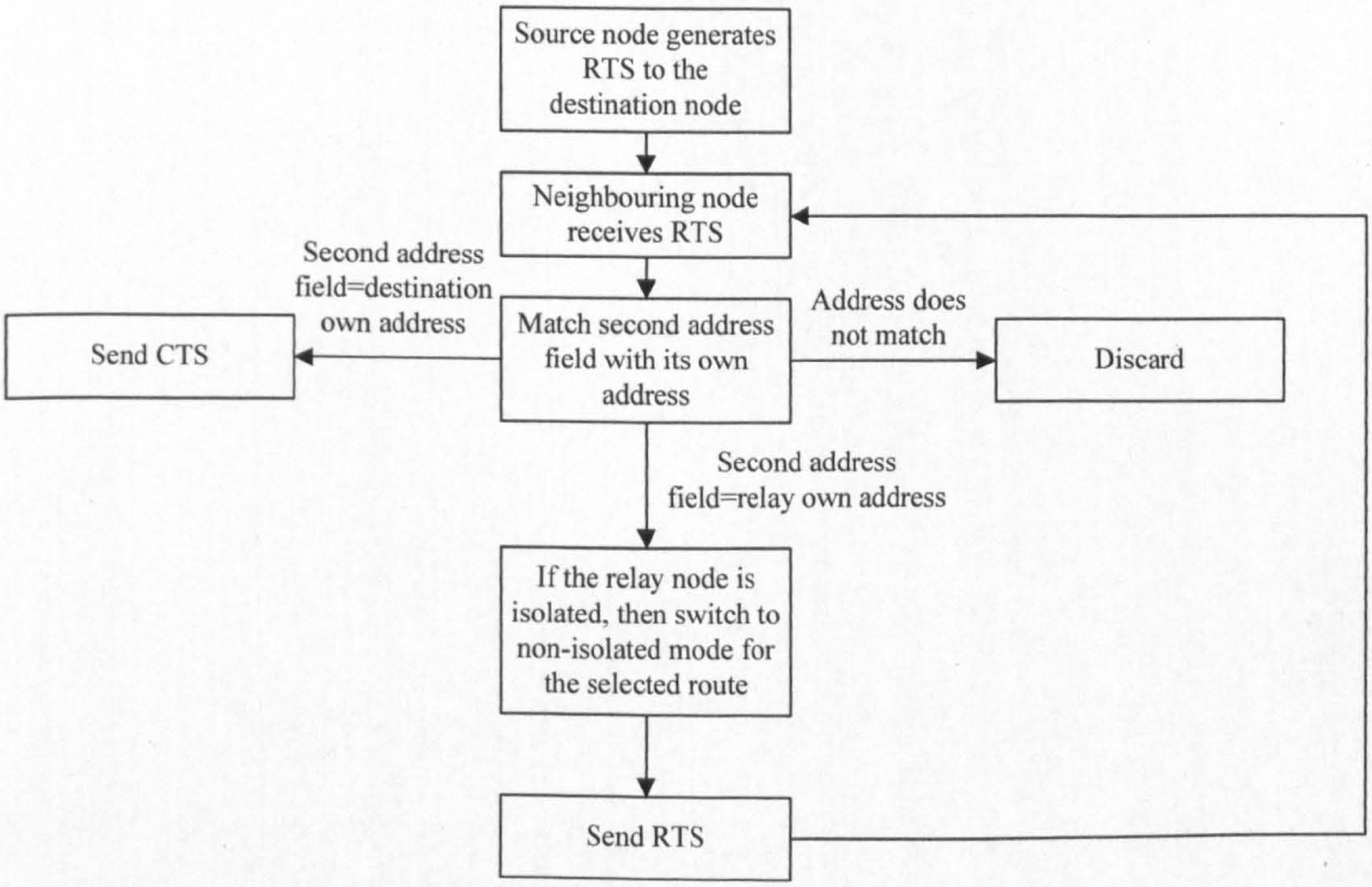


Figure E.2 Flow chart of the RTS/CTS process in an ad hoc network

F Simulation flow (MPR)

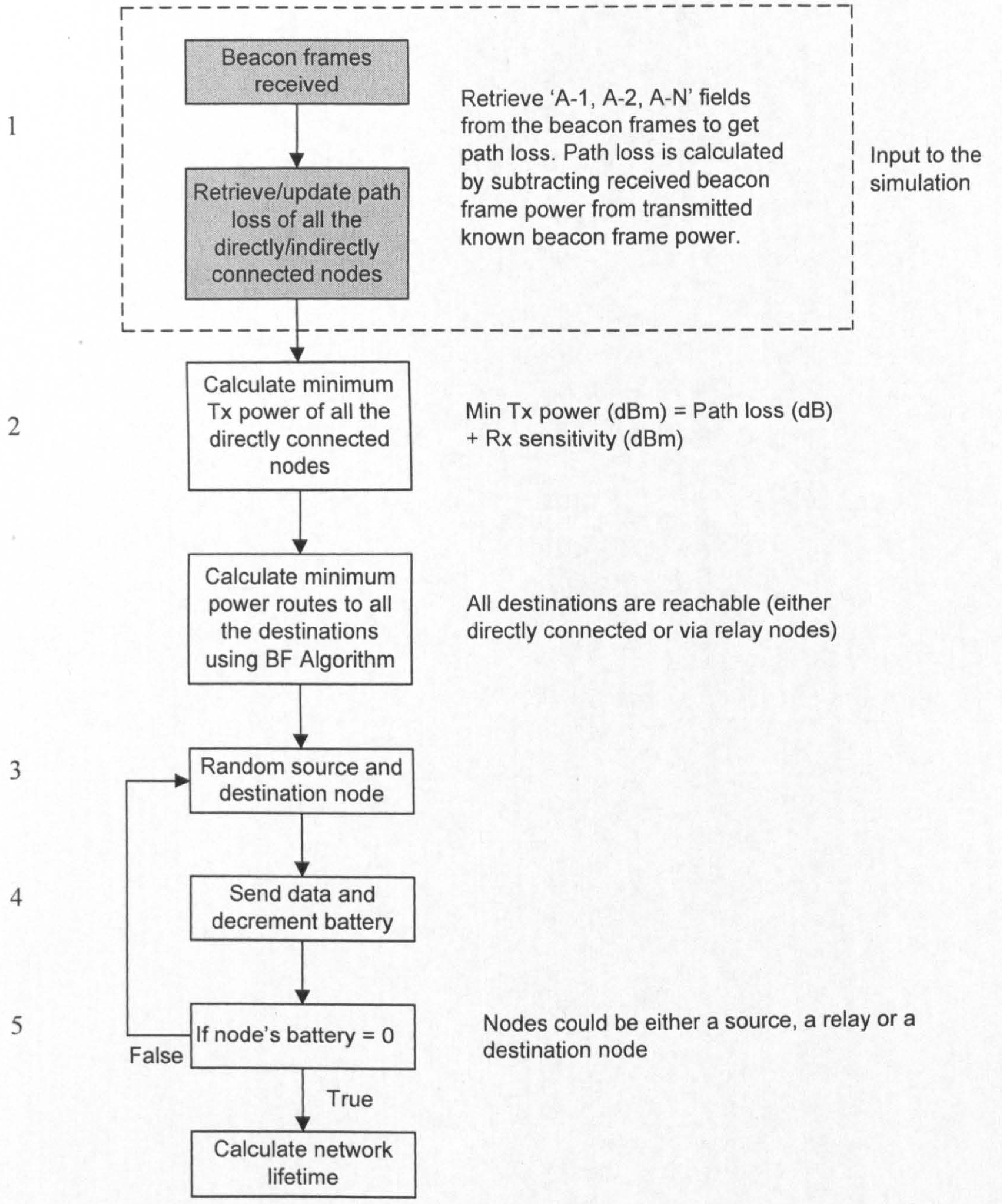


Figure F.1 Simulation flow diagram of MPR approach

Figure F.1 shows the simulation steps when the minimum power approach is used.

1. As discussed in chapter 5, beacon frames are used to keep the network running. Each node is continuously sending and receiving beacon frames in order to share

the necessary information. For example, the "attenuation" field gives the path loss of all the nodes in the network which is used to calculate minimum transmission power route to the destination node.

2. The attenuation information is entered as an input to the simulation and then the minimum power routes for all the directly connected nodes are calculated. Bellman-Ford algorithm is used to calculate minimum power route to the entire destination nodes. Within the set simulation parameters defined in table 6.1, all destination nodes are reachable at the start of the network.
3. Source and destination node are generated randomly and the minimum power route is selected from the routing table.
4. The source node sends 1MB of data and the battery of each node in the route is decremented from the source and the destination nodes. In case of indirectly connected route, battery is also decremented from the relay nodes. The battery decrement process is a function of transmission time, transmission power and processing power.
5. The node's battery is checked after each transmission. The time when the first node's battery reaches end of life, represents the network lifetime.

G Simulation flow (battery charge threshold approach)

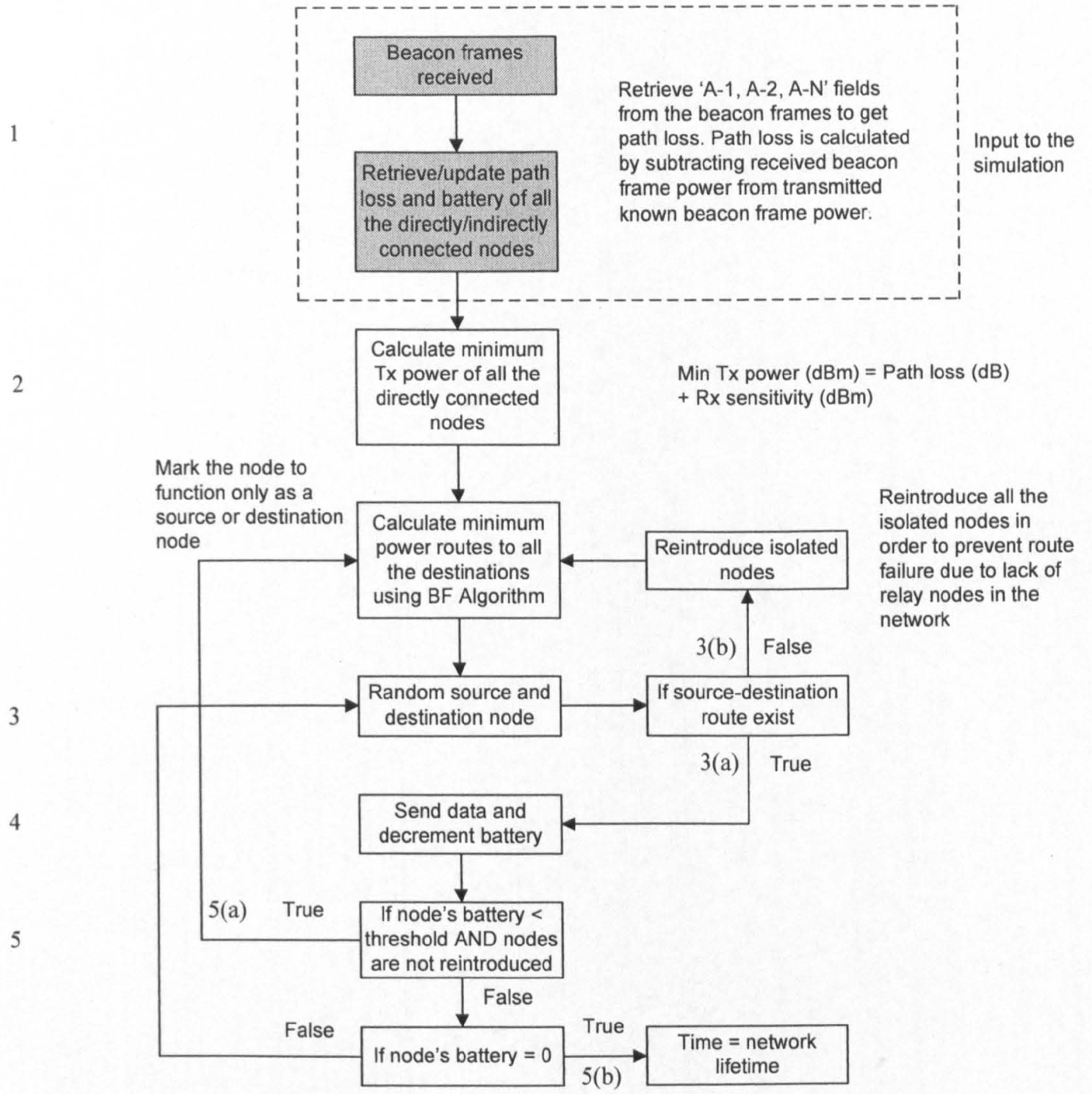


Figure G.1 Simulation flow diagram of battery charge threshold

Figure G.1 shows the simulation steps when a battery charge threshold approach is used.

1. This step is similar to the step 1 (as discussed in Appendix F), however other than attenuation, the battery status information of all the nodes is also taken from the beacon frame.
2. The attenuation information is entered as an input to the simulation and then the minimum power routes for all the directly connected nodes are calculated. Bellman-Ford algorithm is used to calculate minimum power route to the entire

destination nodes. Within the set simulation parameters defined in table 6.1, all destination nodes are reachable at the start of the network.

3. Source and destination nodes are generated randomly. The route selection process selects routes in which all the nodes have remaining battery charge higher than the threshold. If more than one route exists, then the minimum power route is selected. However, there are two possibilities:
    a. If the route exists, then move to step 4.
    b. If the route does not exist, due to a lack of relay nodes in the network. Such a situation arises when a number of nodes have reached threshold and are withdrawn from use as relay nodes. In this situation, all the isolated nodes are reintroduced as relays in the network to prevent route failure.
4. The source node sends 1MB of data and the battery of each node in the route is decremented from the source and the destination nodes. In case of indirectly connected route, battery is also decremented from the relay nodes. The battery decrement process is a function of transmission time, transmission power and processing power.
5. The remaining battery charge of each node that is used as source, relay or destination is checked after each transmission. However an additional check is required.
    a. If node's battery is less than the threshold, then the node is isolated from use as a relay, and the entire routes are recalculated. However, the node can function as a source or a destination node.
    b. If a node's battery reaches zero battery charge, this represent the network lifetime which is defined as the time when the battery charge of the first node in the network is depleted.

H Simulation flow (MPR/MBL approach)

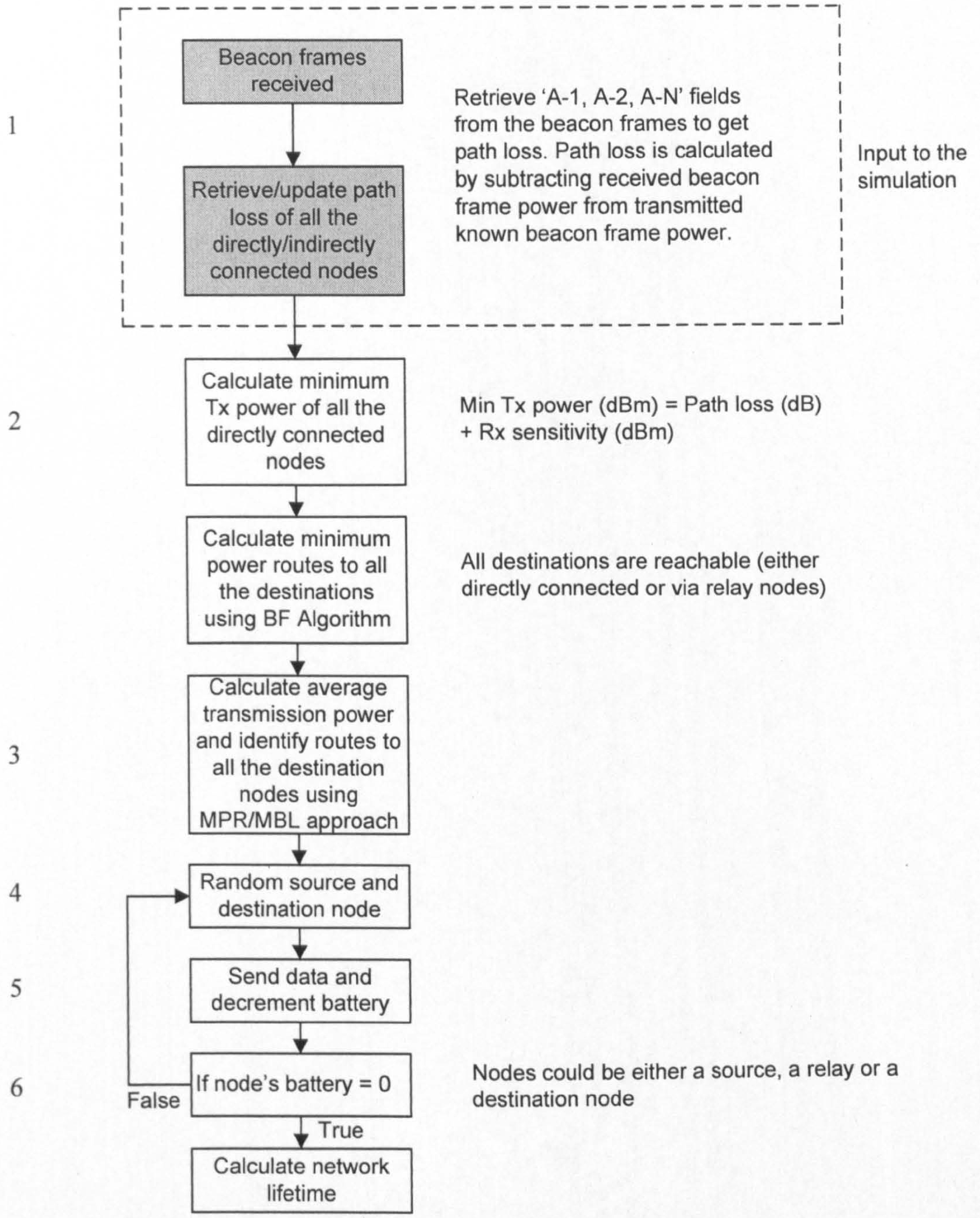


Figure H.1 Simulation flow diagram of MPR/MBL approach

Figure H.1 shows the simulation steps when a MPR/MBL approach is used.

1. This step is similar to the step 1 (as discussed in Appendix F), however other than attenuation, the battery status information of all the nodes is also taken from the beacon frame.

2. The attenuation information is entered as an input to the simulation and then the minimum power routes for all the directly connected nodes are calculated. Bellman-Ford algorithm is used to calculate minimum power route to the entire destination nodes. Within the set simulation parameters defined in table 6.1, all destination nodes are reachable at the start of the network.
3. The average transmission power is calculated with the help of minimum power routes gathered in step 2. The available battery information is gathered in step 1, so routes can be re-calculated by using MPR/MBL approach.
4. Source and destination node are generated randomly and the route is selected using MPR/MBL approach.
5. The source node sends 1MB of data and the battery of each node in the route is decremented from the source and the destination nodes. In case of indirectly connected route, battery is also decremented from the relay nodes. The battery decrement process is a function of transmission time, transmission power and processing power.
6. The node's battery is checked after each transmission. The time when the first node's battery reaches end of life, represents the network lifetime.

I Example route calculation using Bellman-Ford Algorithm

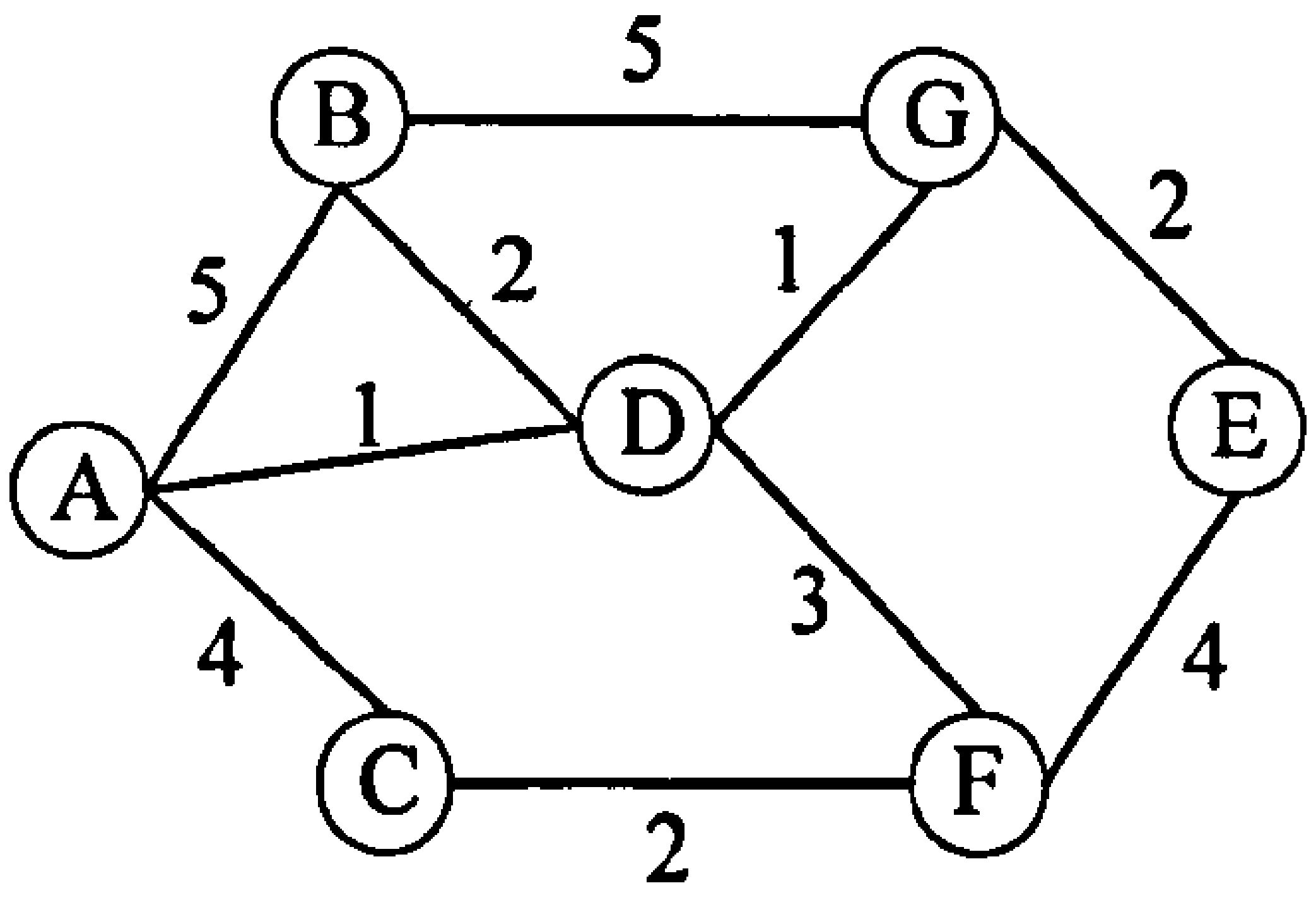


Figure I.1 Seven nodes example network (route calculating using BFA)

D(i,j)[h+1] = min (d(i,k) + D(k,j)[h]) for all i≠j, i=source, j=destination, k neighbours and [h+1]=number of hops

Suppose D(A,E) = min [d(A,k) + D(k,E)], where k = B, C, D

D(i,k) is known because it is directly connected, D(k,j) is not known

In order to calculate D(k,j), first calculate one-hop link (directly connected).

At first iteration (0): D(i,j)[1] = d(i,j)

Node A: D(A,B)[1], D(A,C)[1], D(A,D)[1]

Node B: D(B,A)[1], D(B,D)[1], D(B,G),[1]

Node C: D(C,A)[1], D(C,F)[1]

Node D: D(D,A)[1], D(D,B)[1], D(D,F)][1], D(D,G)[1]

Node E: D(E,F)[1], D(E,G)[1]

Node F: D(F,C)[1], D(F,D)[1], D(F,E)[1]

At second iteration (1):

D(A,B)[2] = min (d(A,B)+D(B,B)[1], d(A,C)+D(C,B)[1], d(A,D)+D(D,B)[1], d(A,E)+D(E,B)[1], d(A,F),+D(F,B)[1])

D(A,B)[2] = min(5+0, 4+∞, 1+2, ∞+∞, ∞+∞) = 3

Nodes which are not reachable either directly (d(i,j)), or at given hop (d(j,k)[h=x]), where x=0,1,2,3..., they are equal to ∞ (no link). After complete execution of enough iteration of all the other destinations, node A topology table looks like Table I.2. Table I.1 is the initial topology table of node A. It can be clearly seen that several destinations are not accessible initially; however after BFA execution such destinations can be accessed.

	A	B	C	D	E	F
A	-	5	4	1	∞	∞
B	5	-	∞	2	∞	∞
C	4	∞	-	∞	∞	2
D	1	2	∞	-	∞	3
E	∞	∞	∞	∞	-	4
F	∞	∞	2	3	4	-

Table I.1 Node A directly connected table

	[0+1]	[1+1]	[2+1]	[3+1]	[4+1]
AB	5	3	3	3	3
AC	4	4	4	4	4
AD	1	1	1	1	1
AE	∞	∞	8	8	8
AF	∞	4	4	4	4

Table I.2 Node A topology table for all the destination

J Allowed power and interference calculation

Consider the following 6-node scenario.

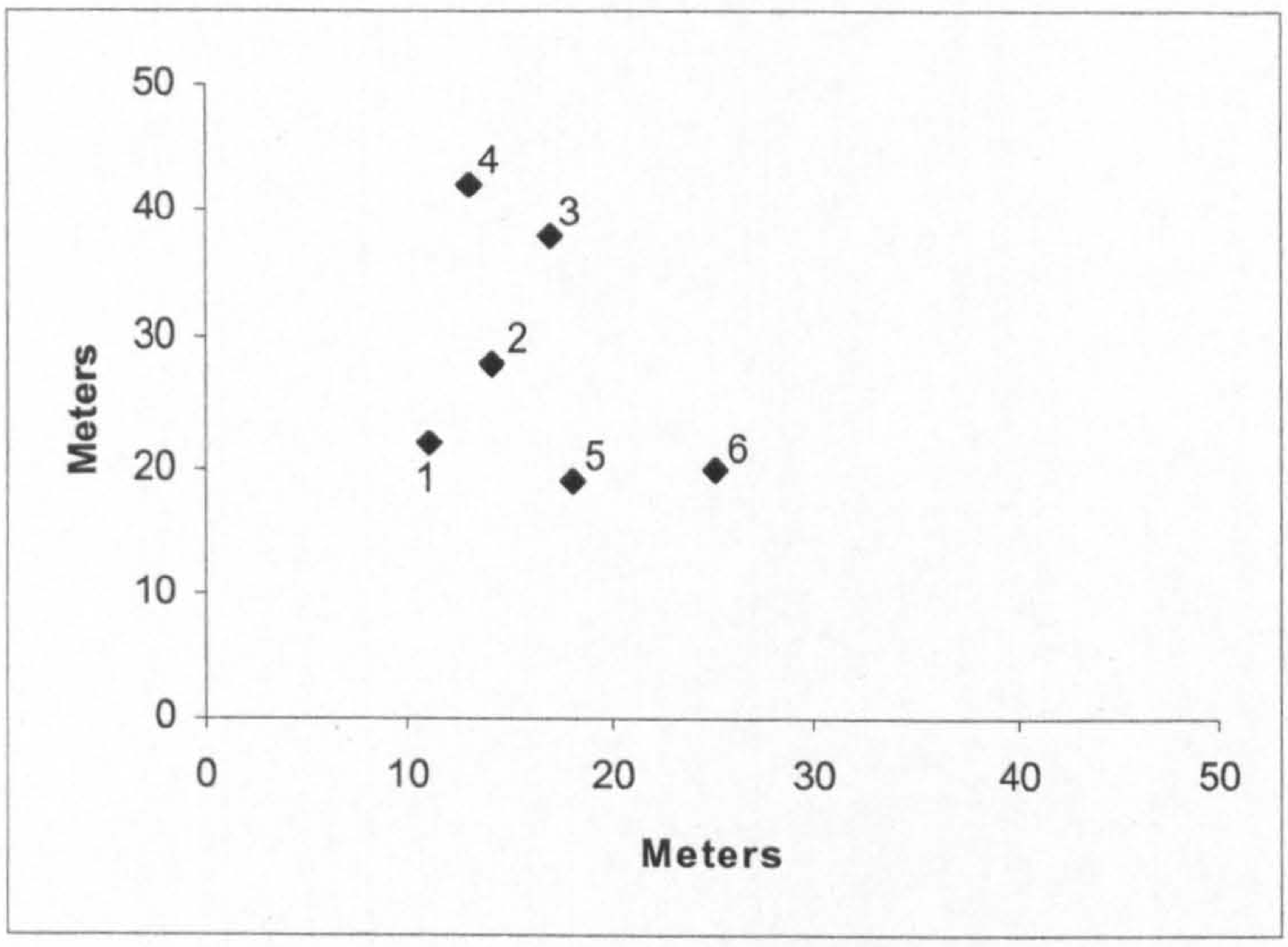


Figure J.1 Six node example scenario

Nodes	Path loss (dB) between the nodes					
	1	2	3	4	5	6
1		83	103	102	75	101
2	83		95	97	91	98
3	103	95		88	105	106
4	102	97	88		106	97
5	75	91	105	106		70
6	101	98	106	97	70	

Table J.1 Path loss (dB) between the nodes

The path loss between the nodes is calculated with the help of equation 5.6 (see section 5.4.1.2, chapter 5).

Link 1-2

Consider that node 1 initiates data transmission with node 2, see Figure J.2. Node 1 knows the path loss between node 2 through beacon frames at the control channel (discussed in section 6.3.1, chapter 6). Node 1 calculates the minimum transmission power to node 2 with the help of the following equation:

Node 1 min Transmission (Tx) power (dBm) = path loss dB (link 1-2) + Rx sensitivity + C/I ratio

Therefore, Node 1 min Tx power (dBm) = 83dB - 86dBm +6dB = +3dBm

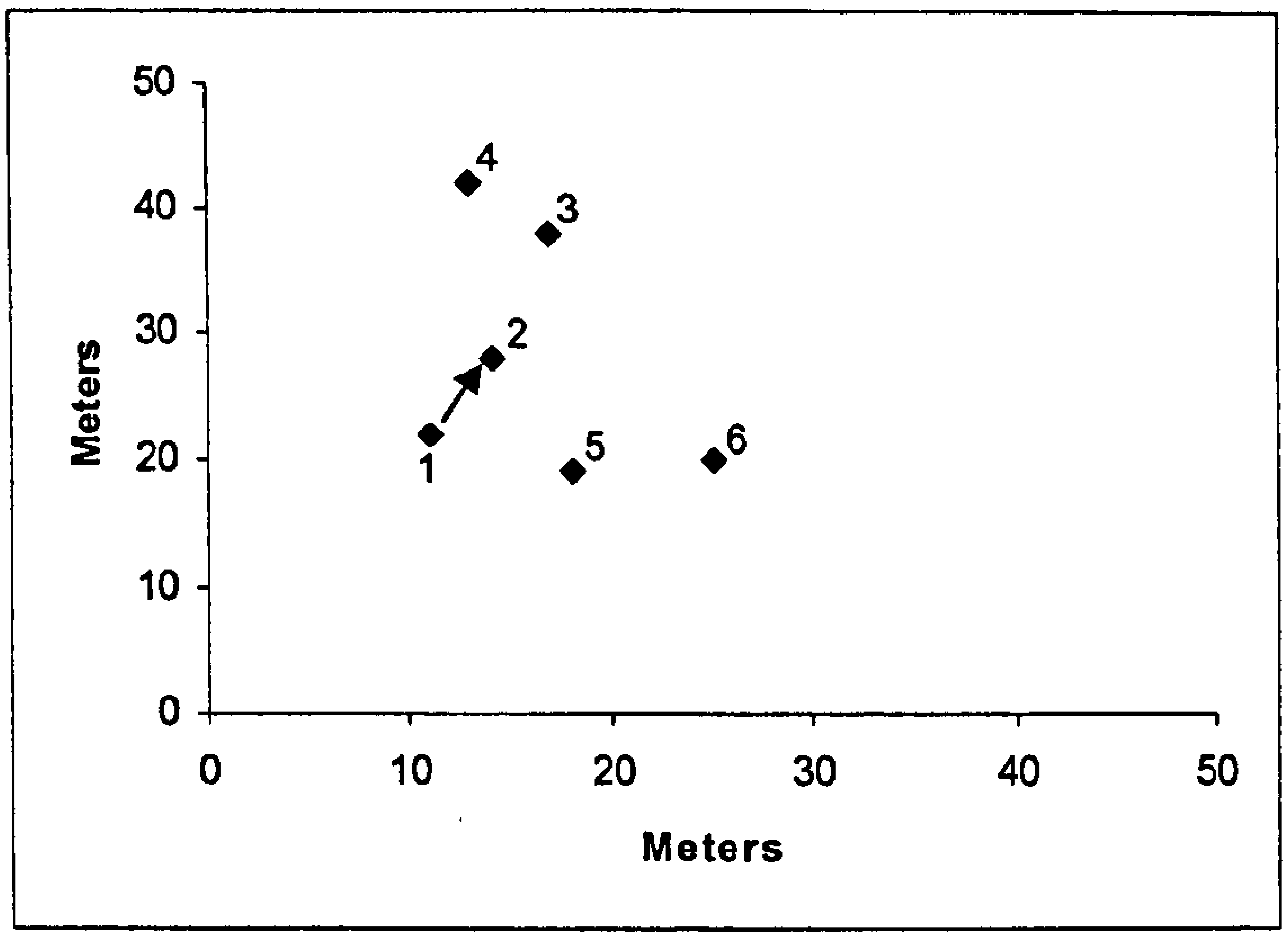


Figure J.2 Data transmission between nodes 1-2

Link 3-4

The transmission of node 1 to node 2 causes interference to the neighbouring nodes. For example, when node 3 initiates data transmission with node 4; node 4 (receiver) can sense signal from the neighbouring transmitter (node 1), see Figure J.3. The interference level at node 4 depends upon the transmission power of node 1 and the path loss between nodes 1 and 4. Hence:

Interference level at node 4 = Tx power (node 1) - path loss (link 1-4) =
+3dBm-102dB = -99dBm

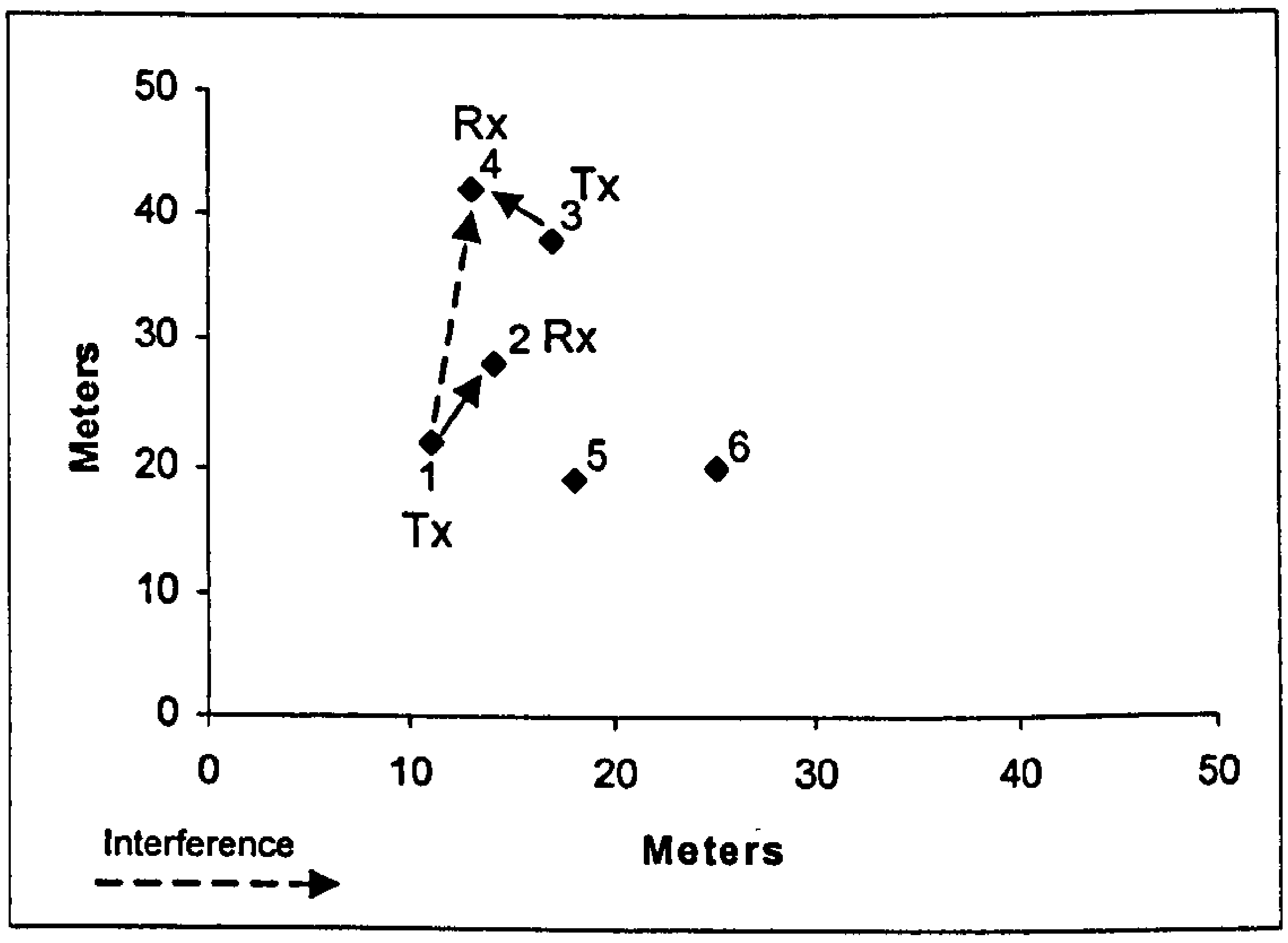


Figure J.3 Data transmission between nodes 3-4

In order to avoid interference with an active receiver (node 2 in this example), the transmission power of node 3 to node 4 should be within the allowed power. The allowed power is set by the most sensitive receiver in the vicinity of the transmitter node. Hence the allowed power for node 3 is set by the receiver of node 2.

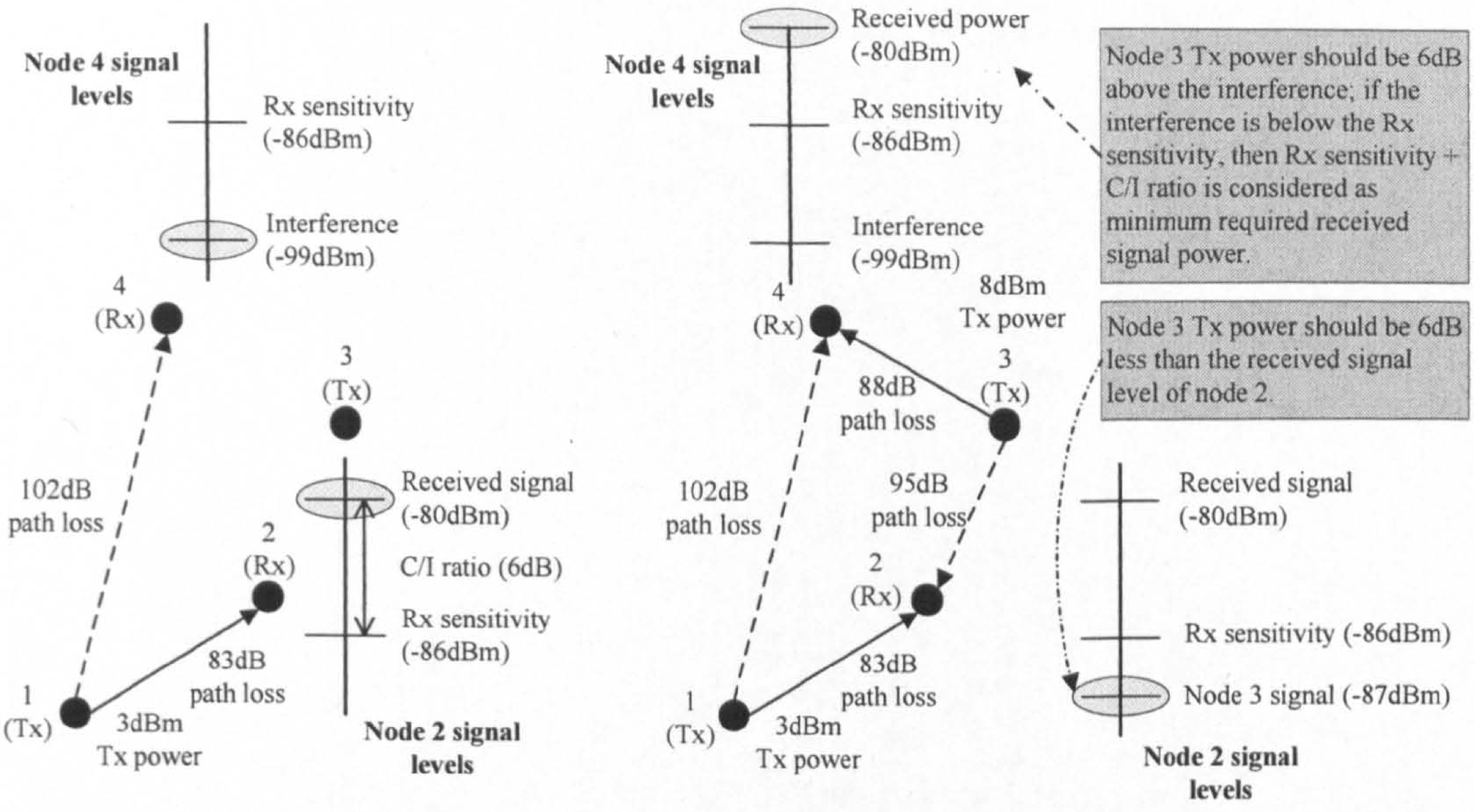
The allowed power for node 3 = path loss dB (link 2-3) + received signal at node 2 - C/I ratio
95dB – 80dBm – 6dB = +9dBm

Node 3 min Tx power (dBm) = path loss dB (link 3-4) + Rx sensitivity + C/I ratio
= 88dB - 86dBm + 6dB = +8dBm

The above Tx power for node 3 is less than the allowed transmission power (+9dBm), hence simultaneous communication between link 3-4 can coexist with link 1-2. There are two conditions for the node 3 Tx power:

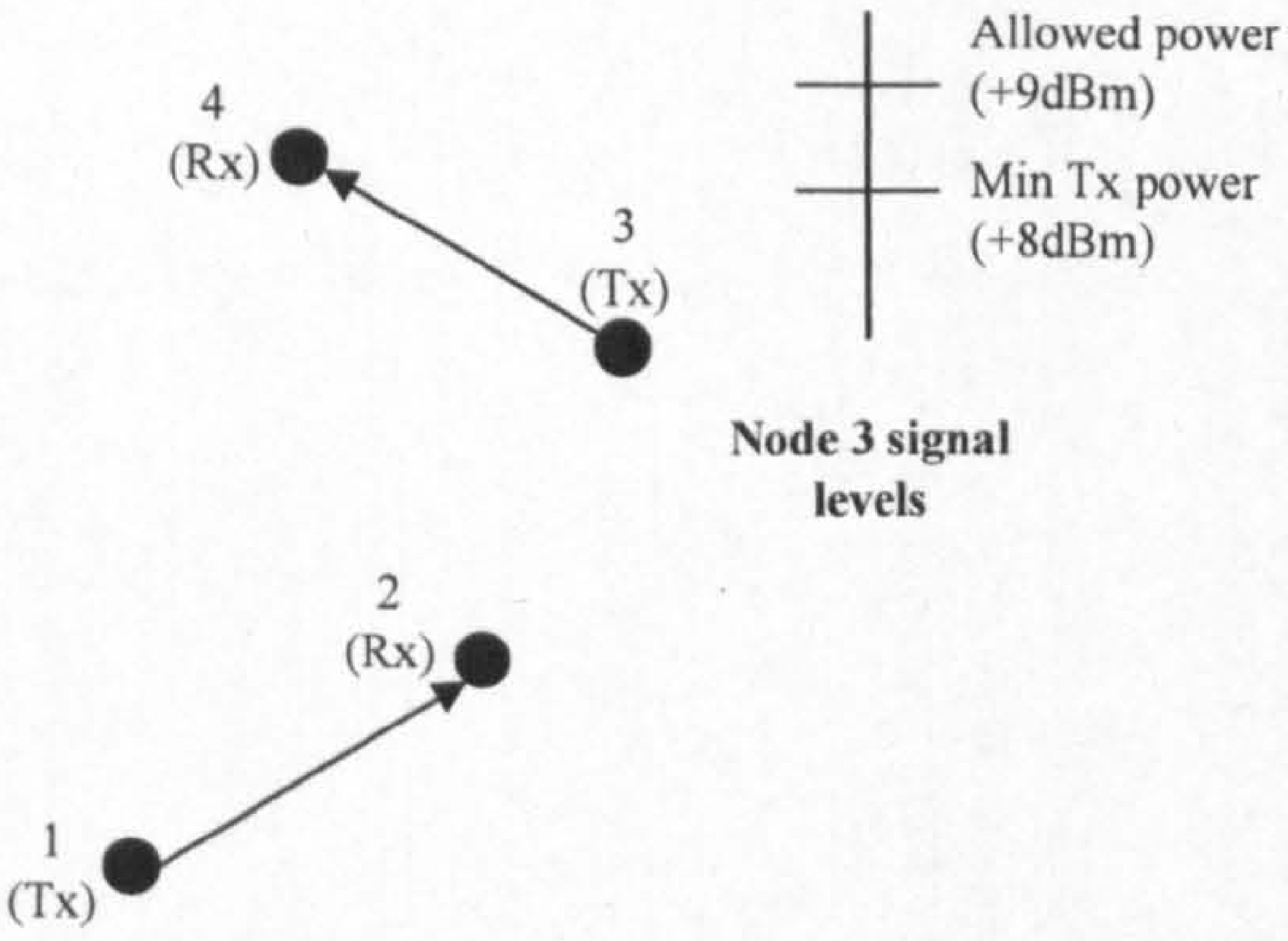
i) Node 3 Tx power must be sufficient to give a power at node 4 that is 6dB above the interference level from node 1, and
ii) Node 3 Tx power must be small enough such that it does not produce interference at node 2.

The above mentioned scheme is illustrated in Figure J.4. Figure J.4(a) shows transmission between link 1-2 and the effect on other neighbouring receivers (node 4 in this example). Figure J.4(b) shows simultaneous transmission between link 3-4 along with link 1-2. It shows that essential conditions (in grey boxes) to be met at the transmitter (node 3) before initiating simultaneous link. Figure J.4(c) shows allowed and minimum power levels of node 3.



a) Node 1-2 transmission

b) Node 3-4 transmission



c) Node 3 power levels

Figure J.4 Minimum, allowed and interference power illustration

Link 6-5

For link 6-5, node 5 is getting interference from the active transmitters in the neighbourhood (node 1 and node 3), see Figure J.5. Therefore monitor the interference from both the transmitters.

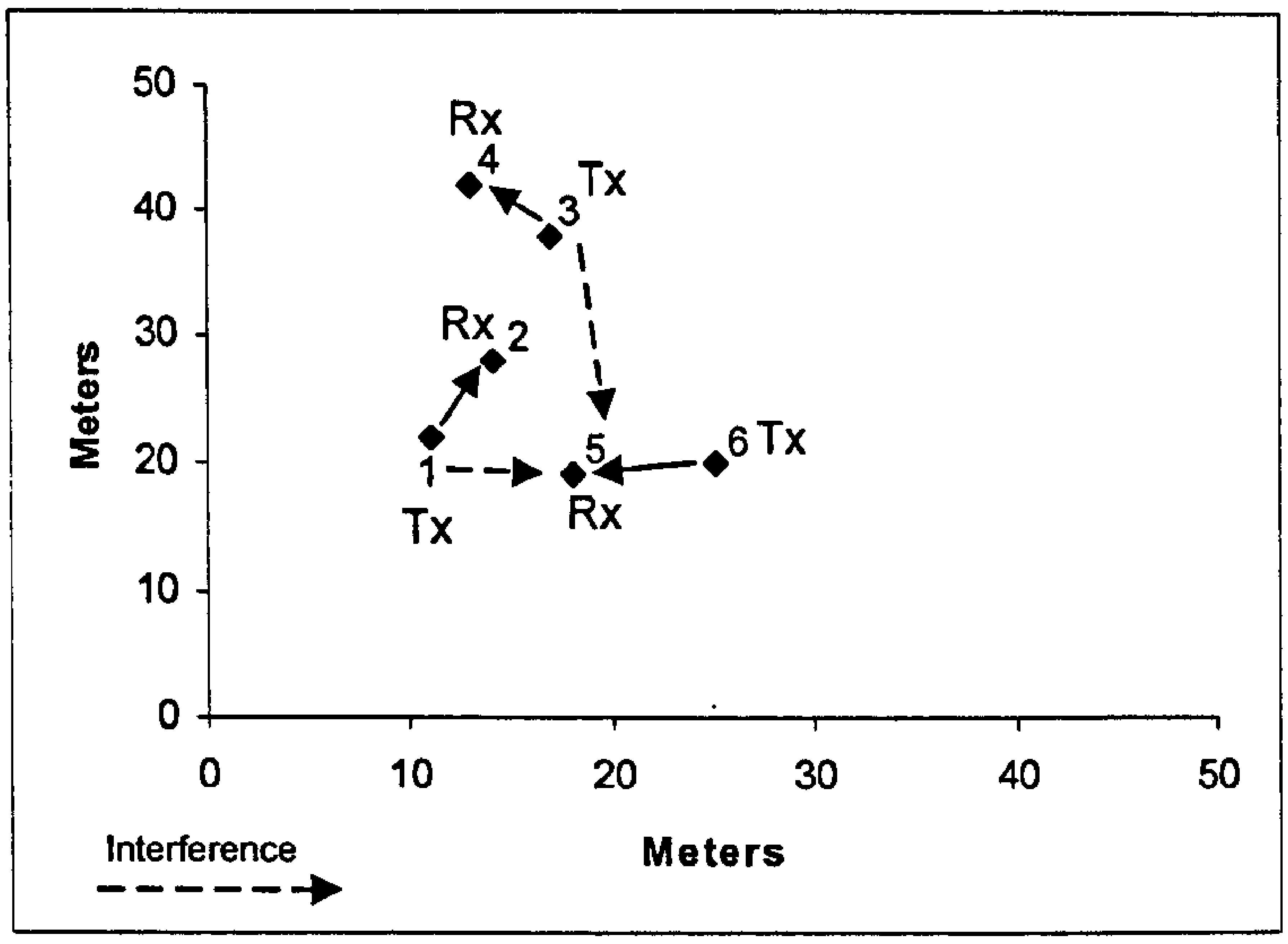


Figure J.5 Data transmission between nodes 6-5

Interference level at node 5 from node 1 = Tx power (node 1) - path loss (link 1-5)

Interference level at node 5 from node 3 = Tx power (node 3) - path loss (link 3-5)

= maximum [(+3dBm - 75dB), (+8dBm - 105dB)] = -72dBm

Hence, node 1 transmission power is creating interference of -72dBm at node 5.

The maximum allowed transmission power for node 6 is set by the level that will not cause a C/I degradation at the neighbouring receivers (node 2 or node 4).

Max allowed power for node 6 with respect to interference at node 2 = received signal at node 2 - C/I + path loss (link 2-6)

Max allowed power for node 6 with respect to interference at node 4 = received signal at node 4 - C/I + path loss (link 4-6)

The maximum allowed power for node 6 is the maximum of these two values = minimum [(-80dBm - 6dB + 98dB), (-80dBm - 6dB + 97dB)] = +11dBm

The interference at node 5 is -72dBm. In order to prevent interference, the received signal power at node 5 should be -66dBm.

Hence:

Node 6 min Tx power (dBm) = Path loss (link 6-5) + received power

= 70dB – 66dBm = +4dBm

The minimum transmission power at node 6 is less than the allowed transmission power, therefore, link 6-5 can co-exist with links 1-2 and 3-4. Figure J.6 illustrates the interference and allowed power levels of node 5 and 6.

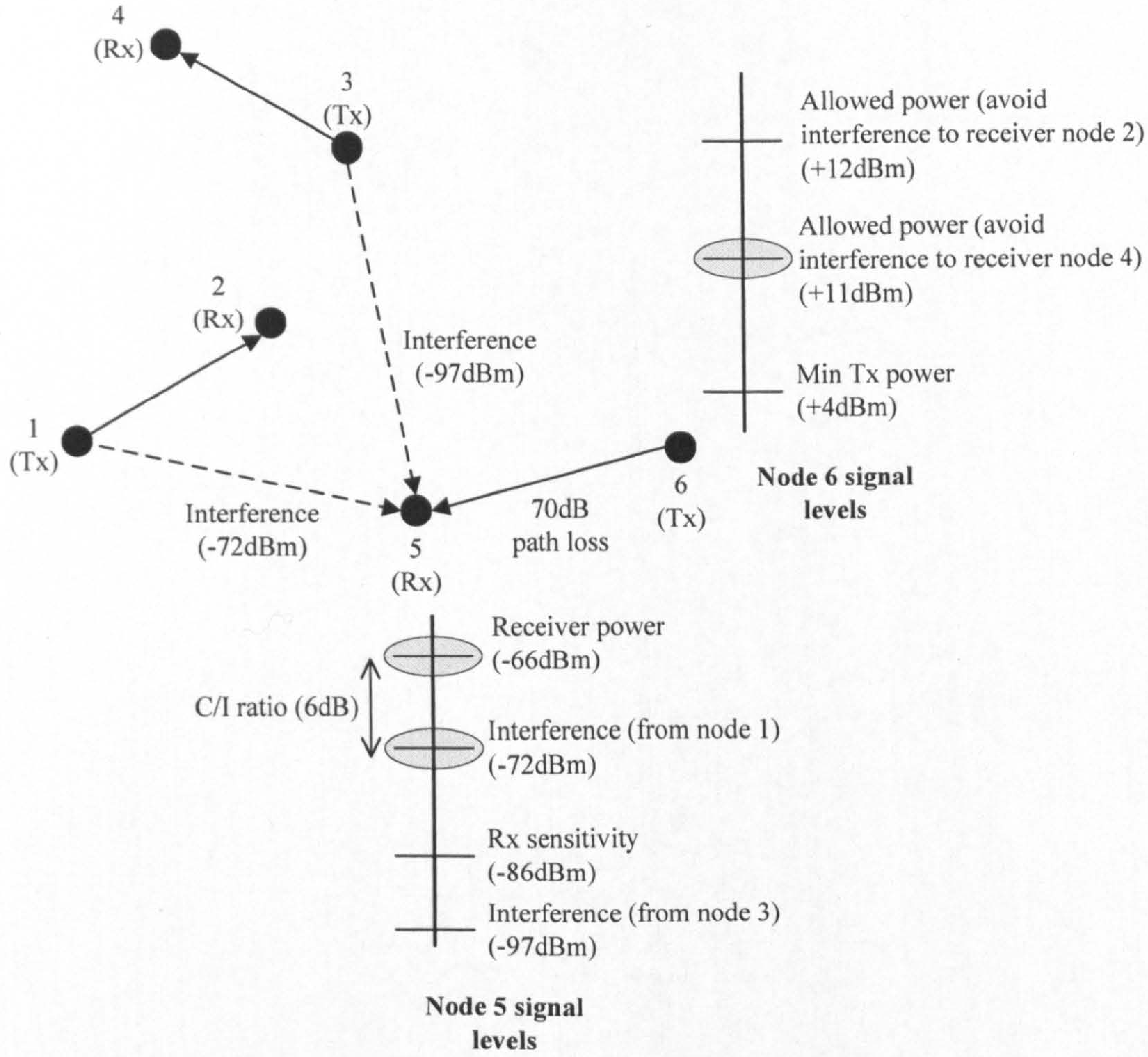


Figure J.6 Node 5 and node 6 signal levels

As discussed in chapter 5, each node broadcasts the maximum power it can receive without causing degradation (receiver power + required C/I ratio) by signalling in the beacon channel. Beacon frames are transmitted at a fixed power known by all the nodes in the network, so the path loss between the nodes can be calculated. When neighbouring nodes receive the maximum received power from all active receivers, they calculate the maximum allowed transmission power.

K Simulation flow – the enhanced design (MPR)

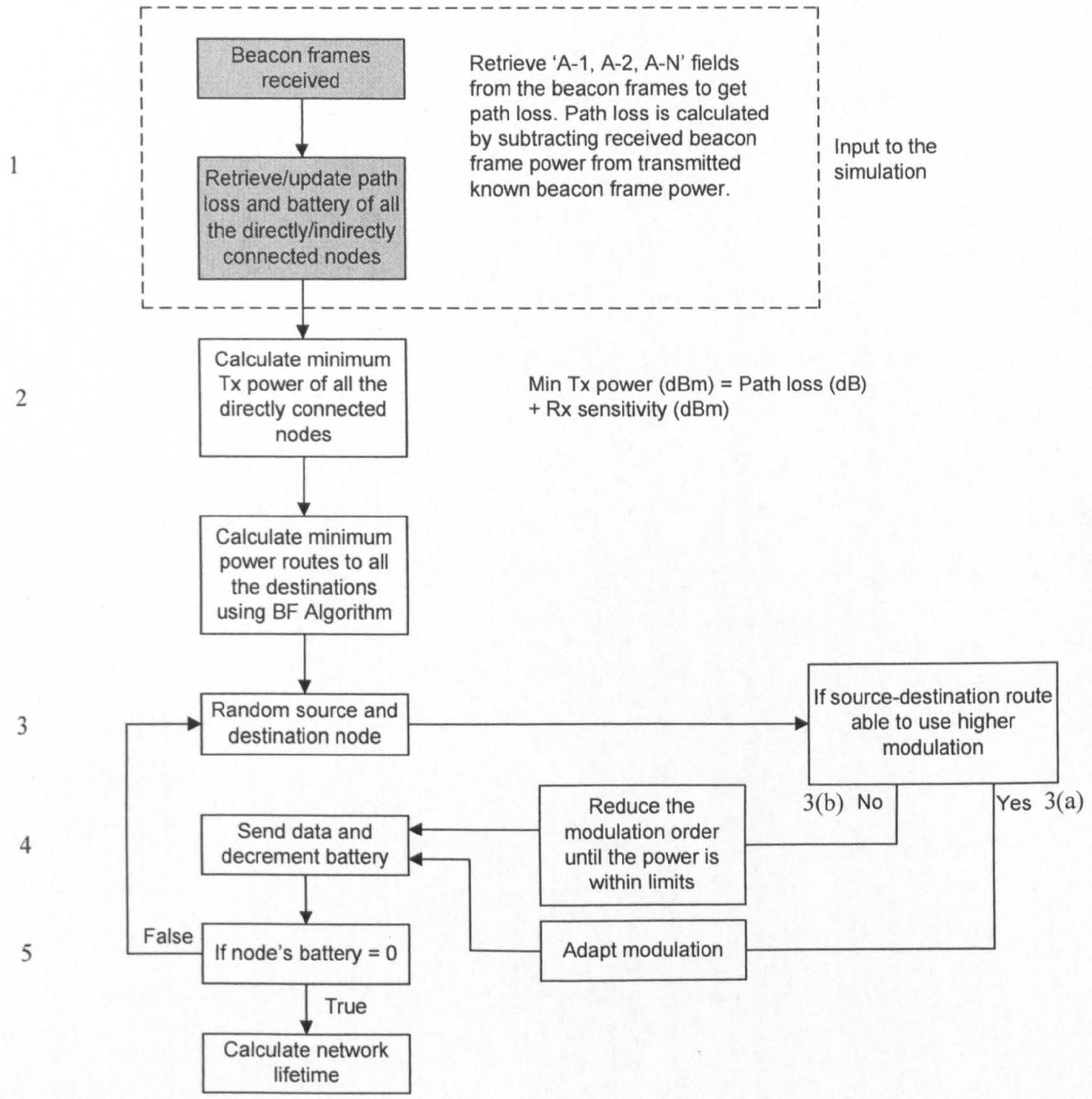


Figure K.1 Simulation flow diagram with MPR and maintaining network throughput per unit time

Figure K.1 shows the simulation steps when throughput is maintained in the multi hop routes using MPR scheme.

1. As discussed in chapter 5, beacon frames are used to broadcast the signalling information. Each node is continuously sending and receiving beacon frames in order to share the necessary information. For example, the "attenuation" field gives the path loss of all the nodes in the network which is used to calculate minimum transmission power route to the destination node. In addition, the battery status information of all the nodes is also taken from the beacon frame.

2. The attenuation information is entered as an input to the simulation and then the minimum power routes for all the directly connected nodes are calculated using the Bellman-Ford algorithm. Using the set of simulation parameters defined in table 6.1, all destination nodes are reachable at the start of the network.
3. Source and destination nodes are generated randomly. The minimum power route is selected, however there are two possibilities:
    a. Modulation is adapted to maintain throughput per unit time for the minimum power route, and then power is adjusted according to the selected modulation scheme. The source node checks that the required transmission power is less than the maximum allowed limit of 100mW.
    b. If there is insufficient margin in the power budget to allow the required higher order modulation for the selected route, then the source node reduces the modulation order until the power is within limits with a consequent compromise in the throughput per unit time.
4. The source node sends 1MB of data and the battery of each node in the route is decremented from the source and the destination nodes. In the case of an indirectly connected route, the battery status of the relay nodes is also updated. The battery usage is a function of the transmission time, transmission power and processing power.
5. The node's battery is checked after each transmission. The time when the first node's battery reaches end of life, this represents the network lifetime.

L Simulation flow – the enhanced design (battery charge threshold)

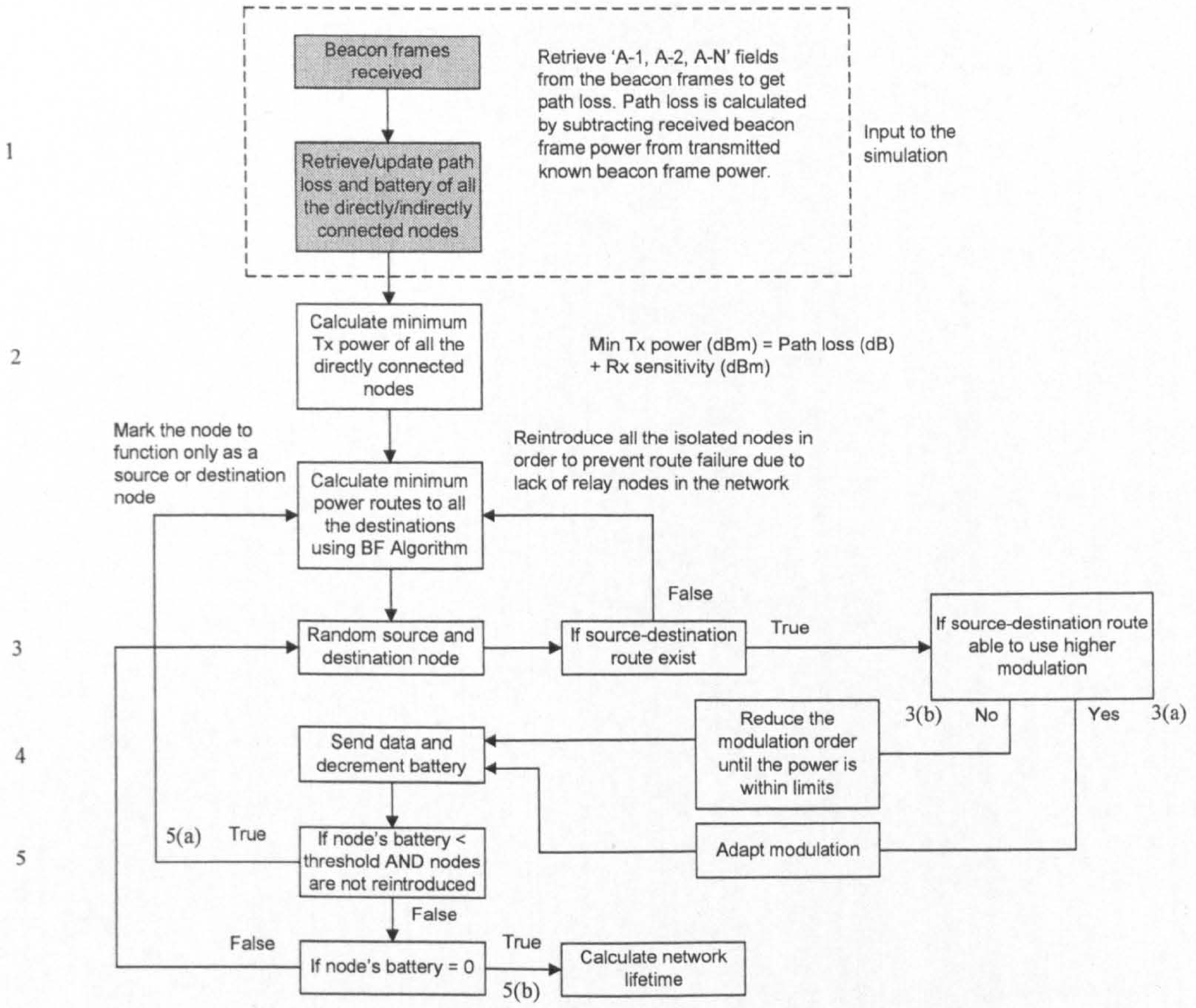


Figure L.1 Simulation flow diagram with battery threshold and maintaining network throughput per unit time

Figure L.1 shows the simulation steps when throughput is maintained in the multi hop routes using battery charge threshold scheme.

1. As discussed in chapter 5, beacon frames are used to broadcast the signalling information. Each node is continuously sending and receiving beacon frames in order to share the necessary information. For example, the "attenuation" field gives the path loss of all the nodes in the network which is used to calculate minimum transmission power route to the destination node. In addition, the battery status information of all the nodes is also taken from the beacon frame.
2. Attenuation information is entered as an input to the simulation and then the minimum power routes for all the directly connected nodes are calculated using

the Bellman-Ford algorithm. Using the set of simulation parameters defined in table 6.1, all destination nodes are reachable at the start of the network.

3. Source and destination nodes are generated randomly. The route selection process identifies routes in which all the nodes have available battery higher than threshold. The minimum power route is selected, however, there are two possibilities:
    a. Modulation is adapted to maintain throughput per unit time for the minimum power route, and then power is adjusted according to the selected modulation scheme. The source node checks that the required transmission power is less than the maximum allowed limit of 100mW.
    b. If there is insufficient margin in the power budget to allow the required higher order modulation for the selected route, then the source node reduces the modulation order until the power is within limits with a consequent compromise in the throughput per unit time.
4. The source node sends 1MB of data and the battery of each node in the route is decremented from the source and the destination nodes. In the case of an indirectly connected route, the battery status of the relay nodes is also updated. The battery usage is a function of the transmission time, transmission power and processing power.
5. The remaining battery charge of each node that is used as source, relay or destination is checked after each transmission. However an additional check is required.
    a. If node's battery is less than the threshold, then the node is isolated from use as a relay, and the entire routes are recalculated. However, the node can function as a source or a destination node.
    b. If a node's battery reaches zero battery charge, this represent the network lifetime which is defined as the time when the battery charge of the first node in the network is depleted.

M Simulation flow – the enhanced design (MPR/MBL)

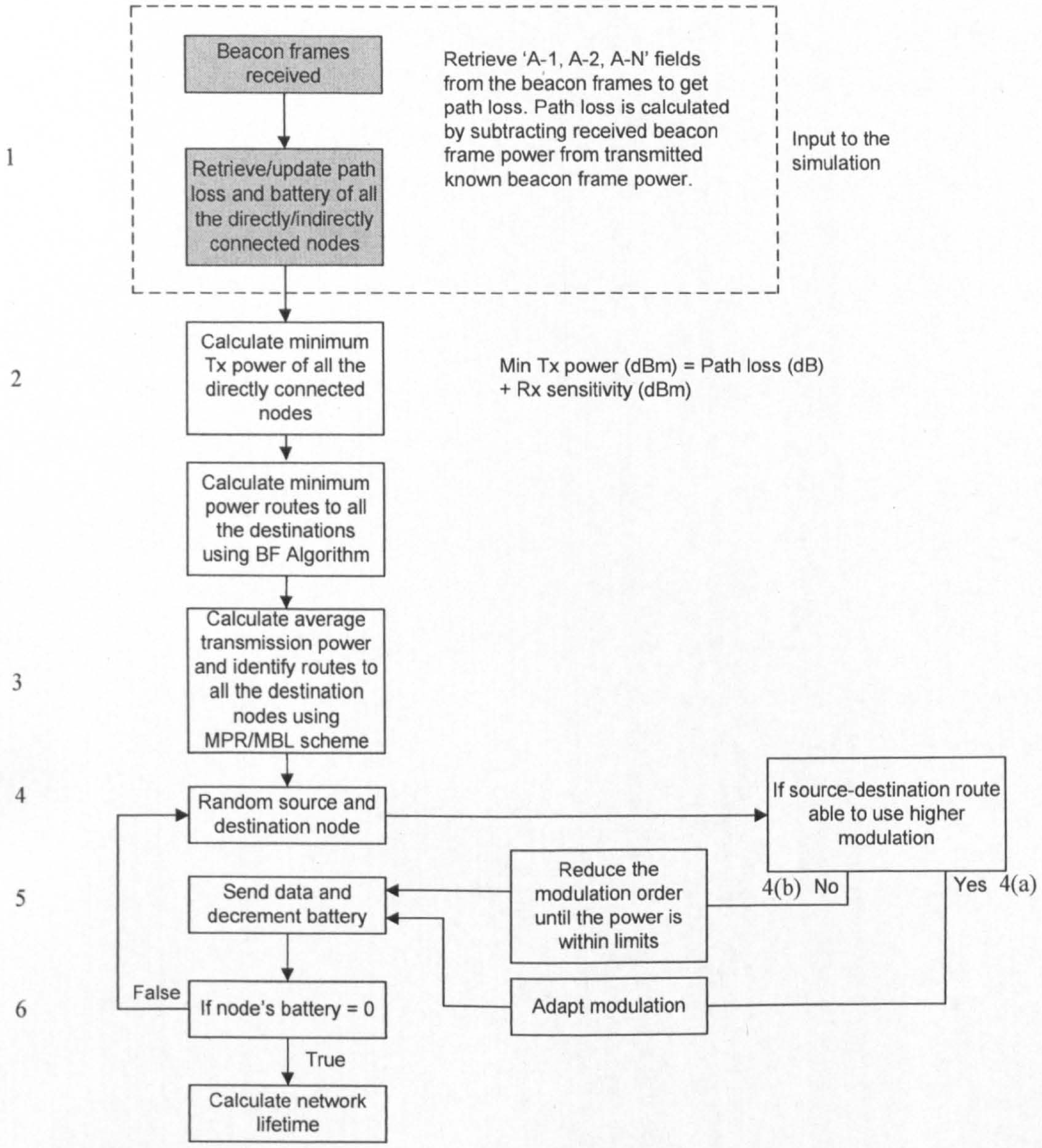


Figure M.1 Simulation flow diagram with MPR/MBL scheme and maintaining network throughput per unit time

Figure M.1 shows the simulation steps when throughput is maintained in the multi hop routes using MPR/MBL scheme.

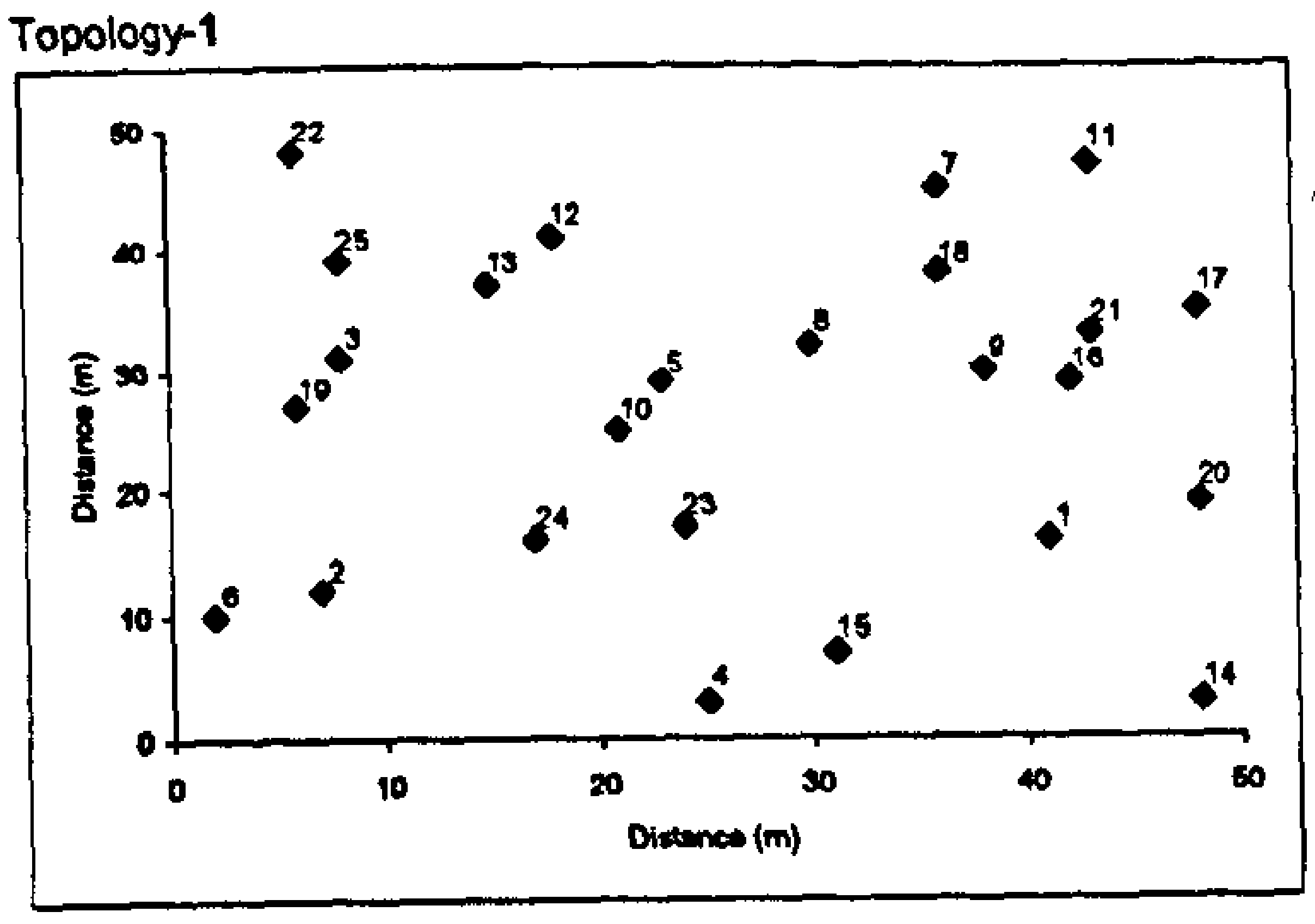
1. As discussed in chapter 5, beacon frames are used to broadcast the signalling information. Each node is continuously sending and receiving beacon frames in order to share the necessary information. For example, the "attenuation" field gives the path loss of all the nodes in the network which is used to calculate

minimum transmission power route to the destination node. In addition, the battery status information of all the nodes is also taken from the beacon frame.

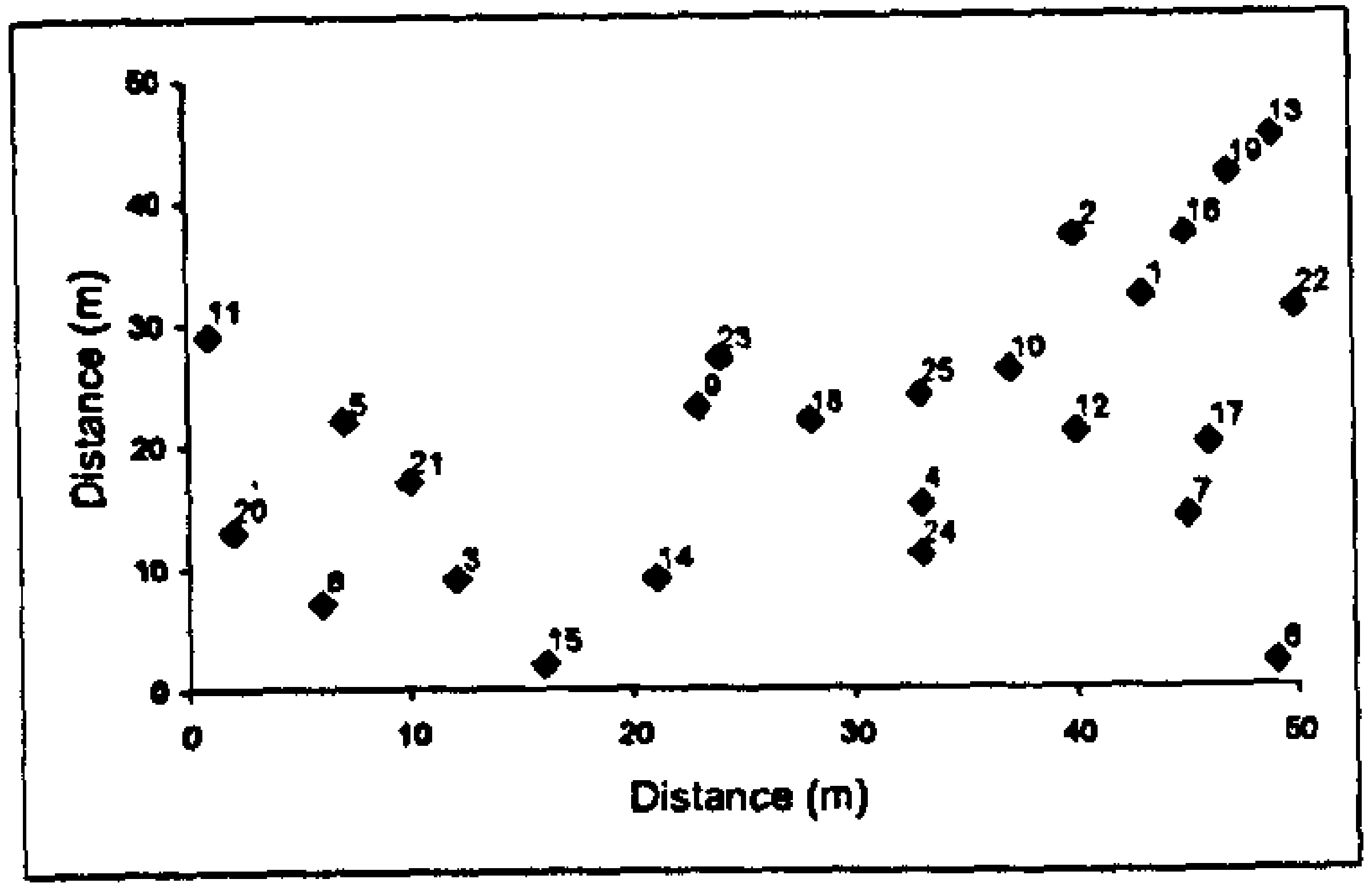
2. The attenuation information is entered as an input to the simulation and then the minimum power routes for all the directly connected nodes are calculated using the Bellman-Ford algorithm. Using the set of simulation parameters defined in table 6.1, all destination nodes are reachable at the start of the network.
3. The average transmission is calculated with the help of minimum power routes gathered in step 2. The available battery information is gathered in step 1, so route can be re-calculated by using MPR/MBL scheme.
4. Source and destination nodes are generated randomly and the route is selected using MPR/MBL scheme. However there are two possibilities:
    a. Modulation is adapted to maintain throughput per unit time for the given route, and then power is adjusted according to the selected modulation scheme. The source node checks that the required transmission power is less than the maximum allowed limit of 100mW.
    b. If there is insufficient margin in the power budget to allow the required higher order modulation for the selected route, then the source node reduces the modulation order until the power is within limits with a consequent compromise in the throughput per unit time.
5. The source node sends 1MB of data and the battery of each node in the route is decremented from the source and the destination nodes. In the case of an indirectly connected route, the battery status of the relay nodes is also updated. The battery usage is a function of the transmission time, transmission power and processing power.
6. The node's battery is checked after each transmission. The time when the first node's battery reaches end of life, this represents the network lifetime.

N Different network topologies

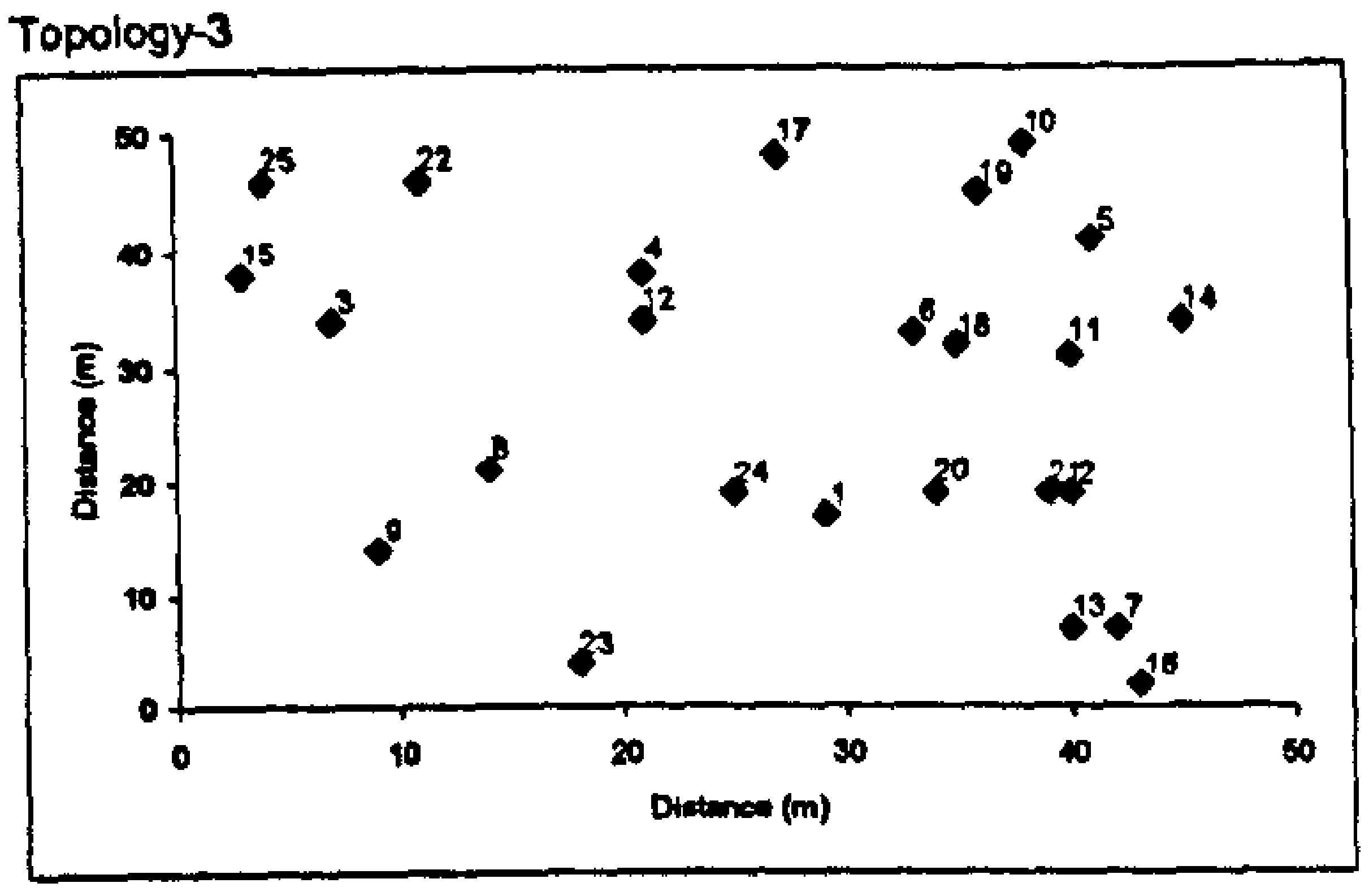
Topology-1



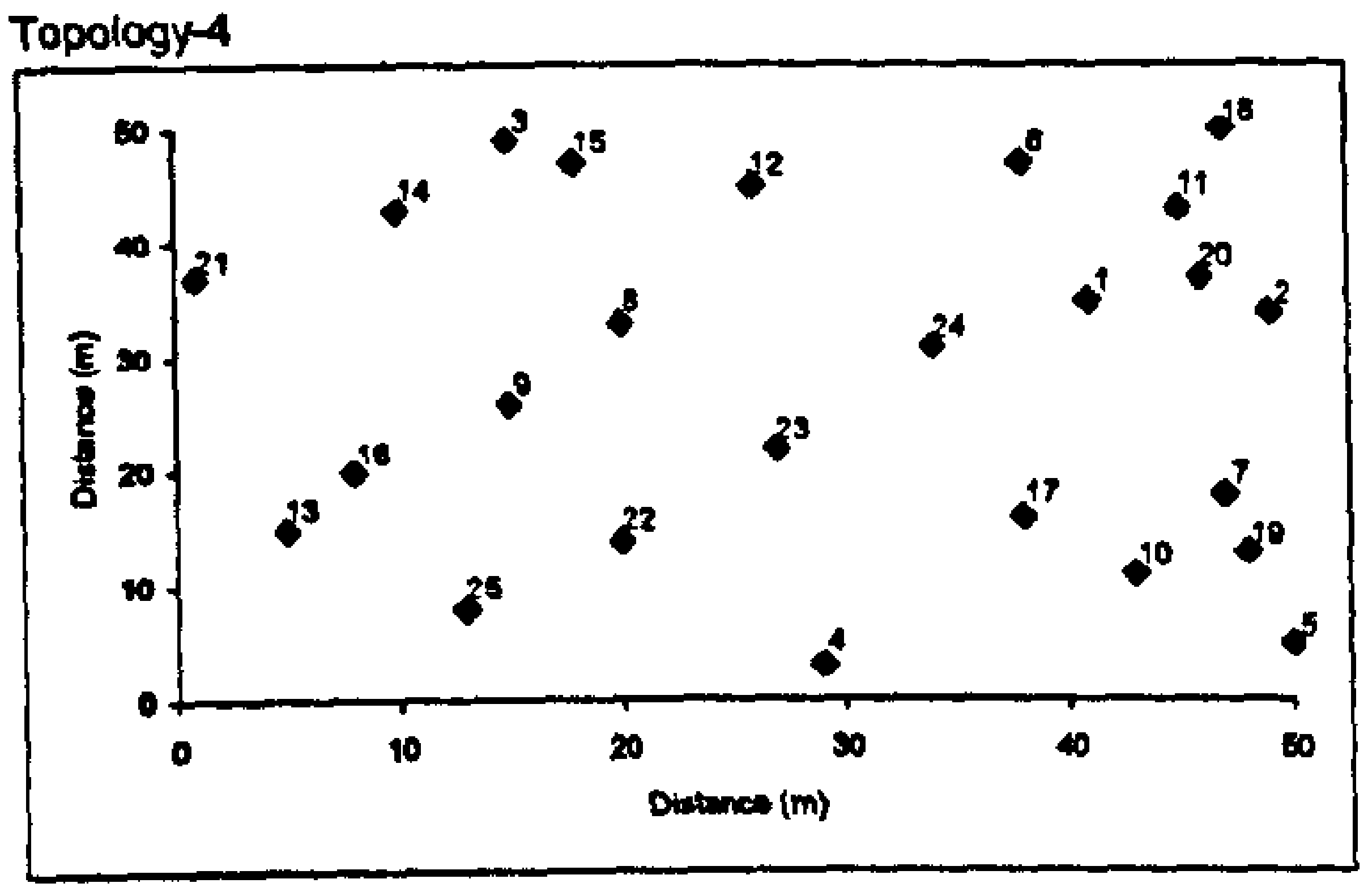
Topology-2



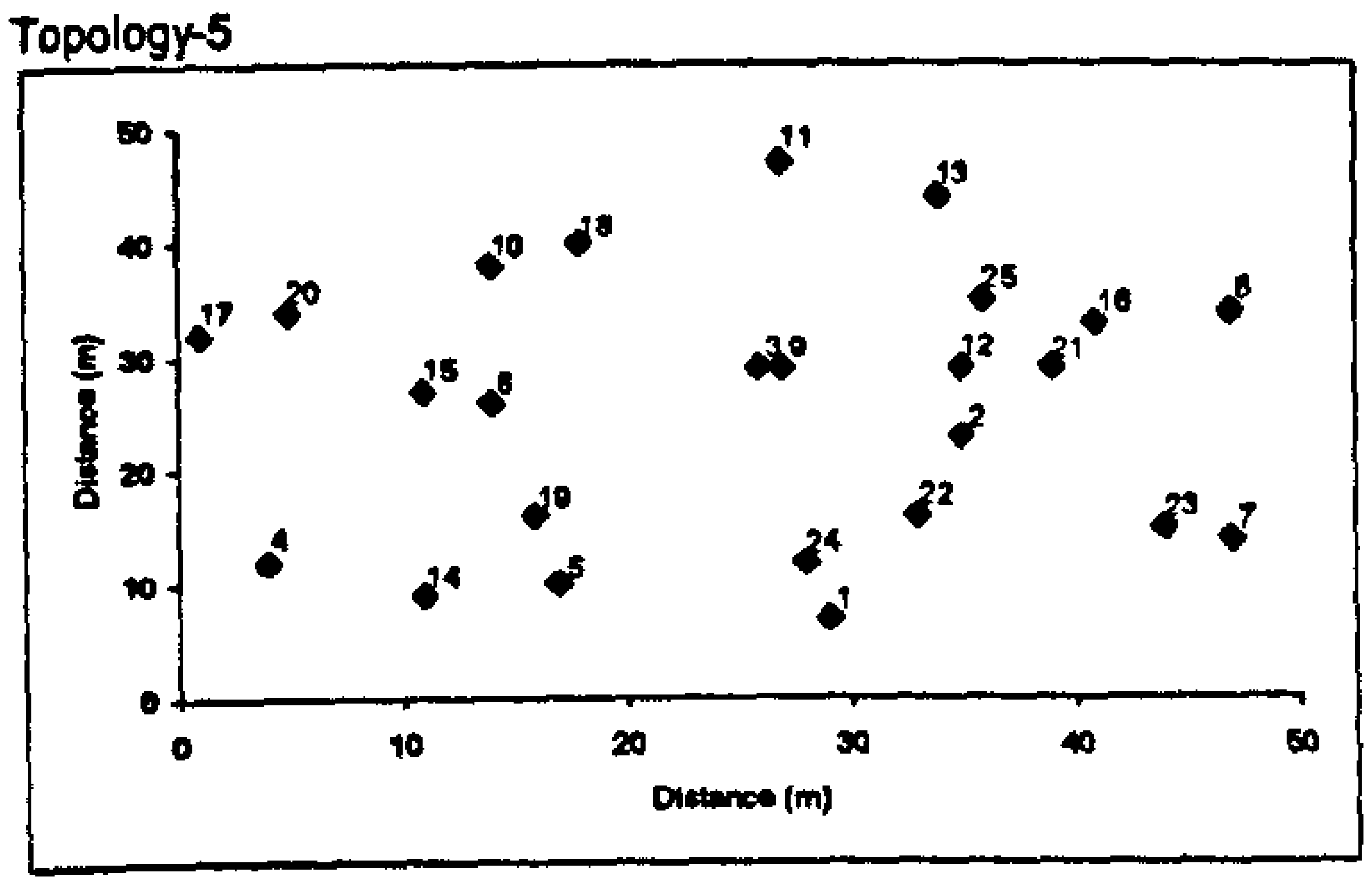
Topology-3



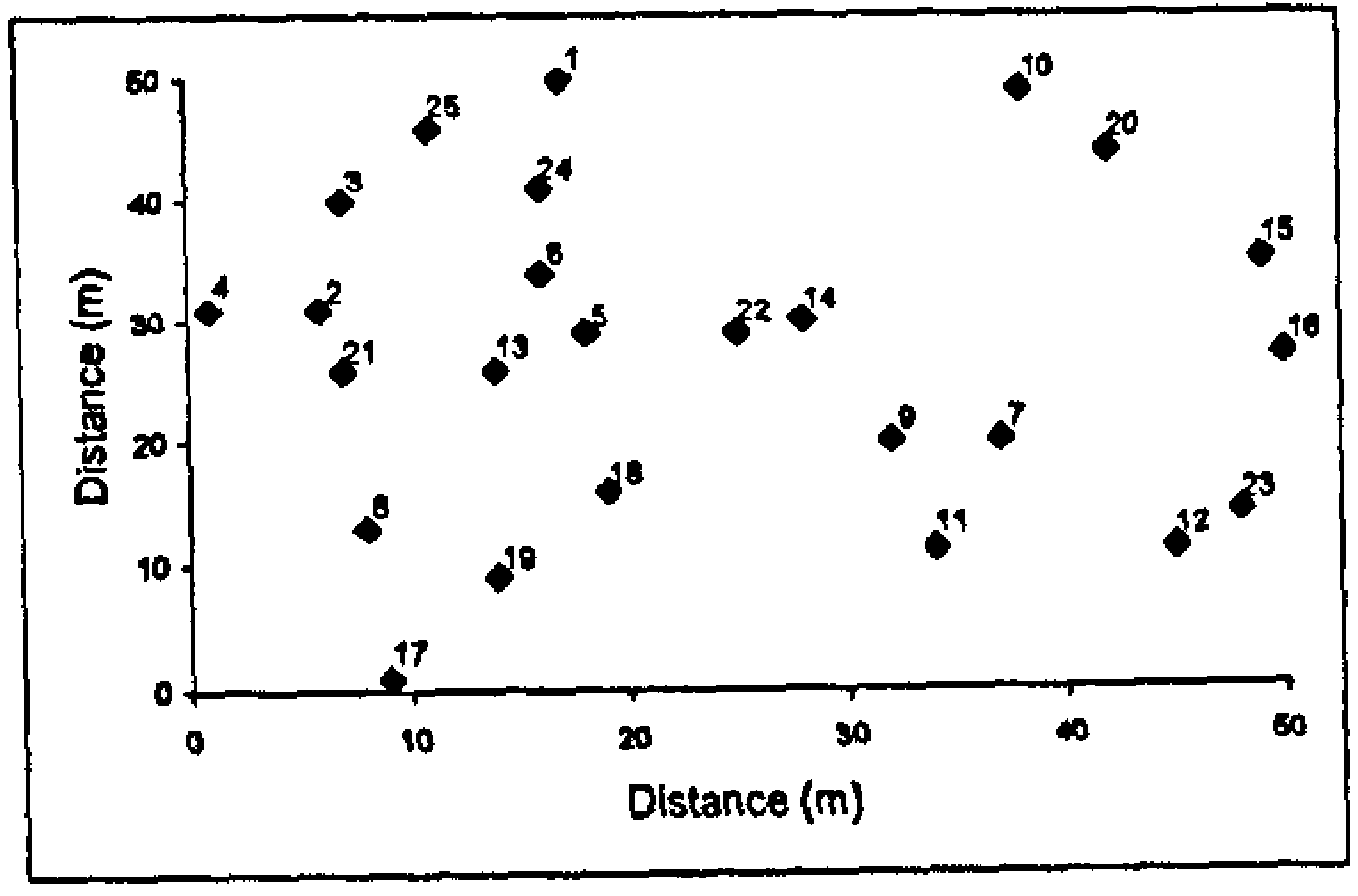
Topology-4



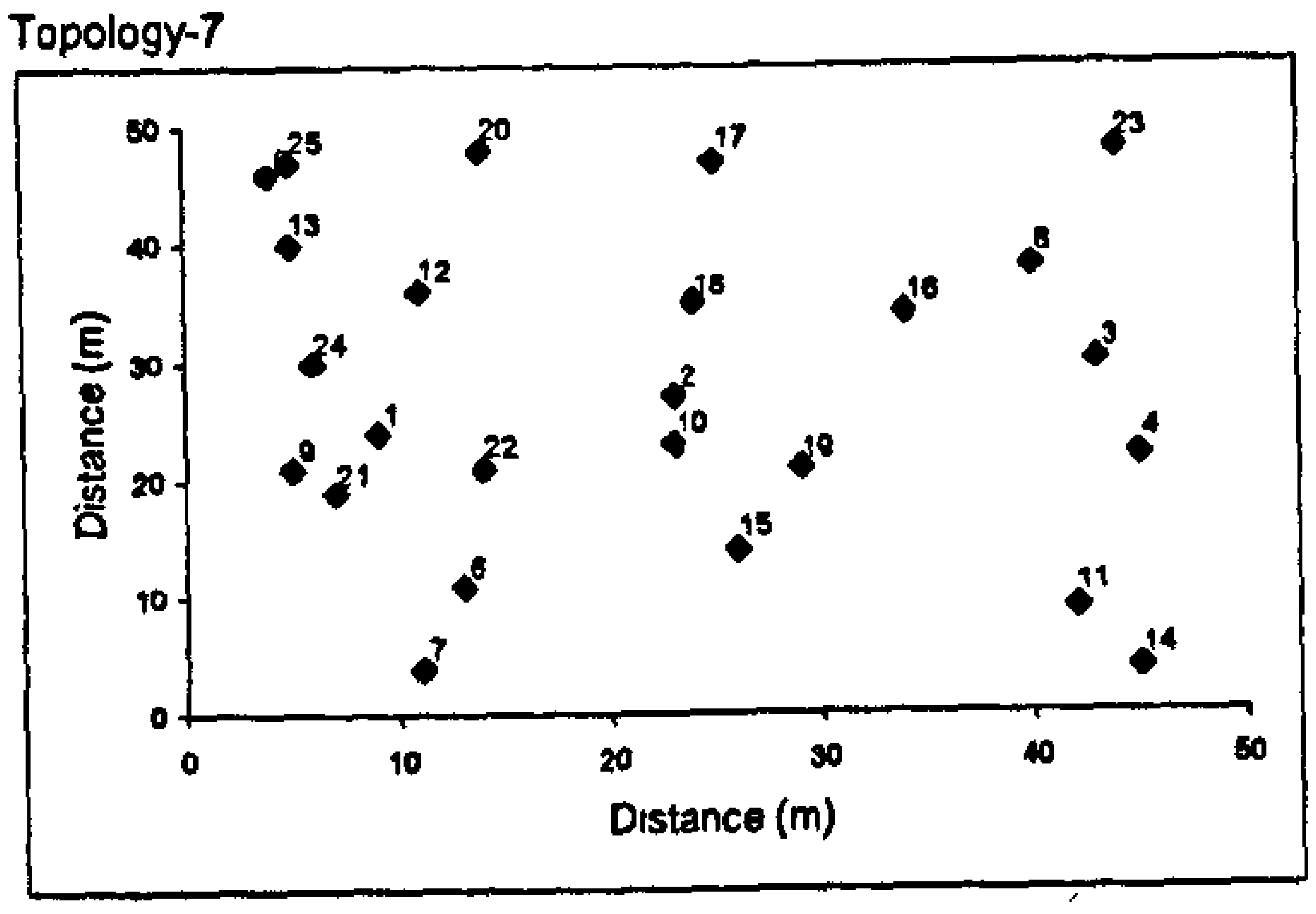
Topology-5



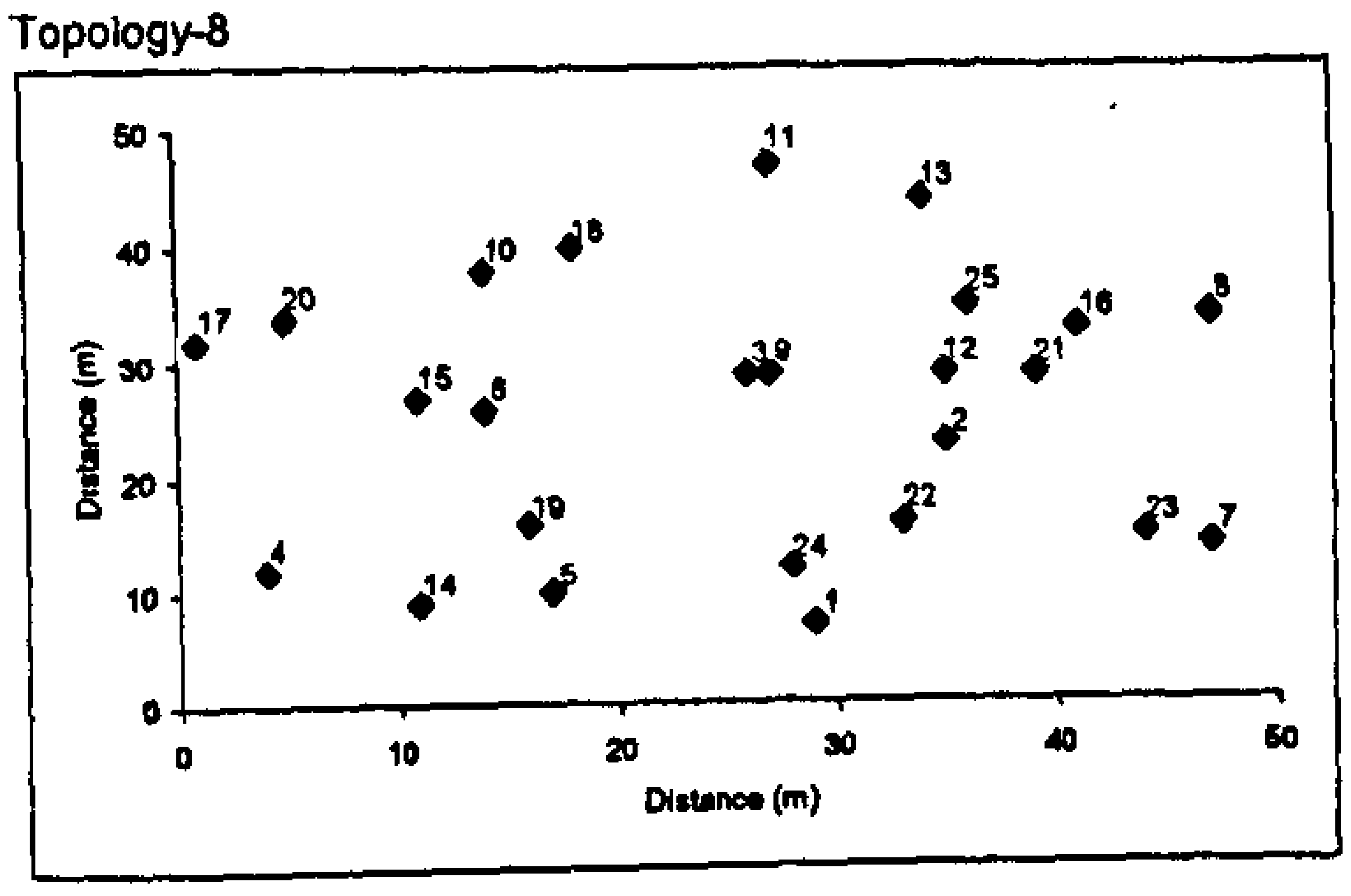
Topology-6



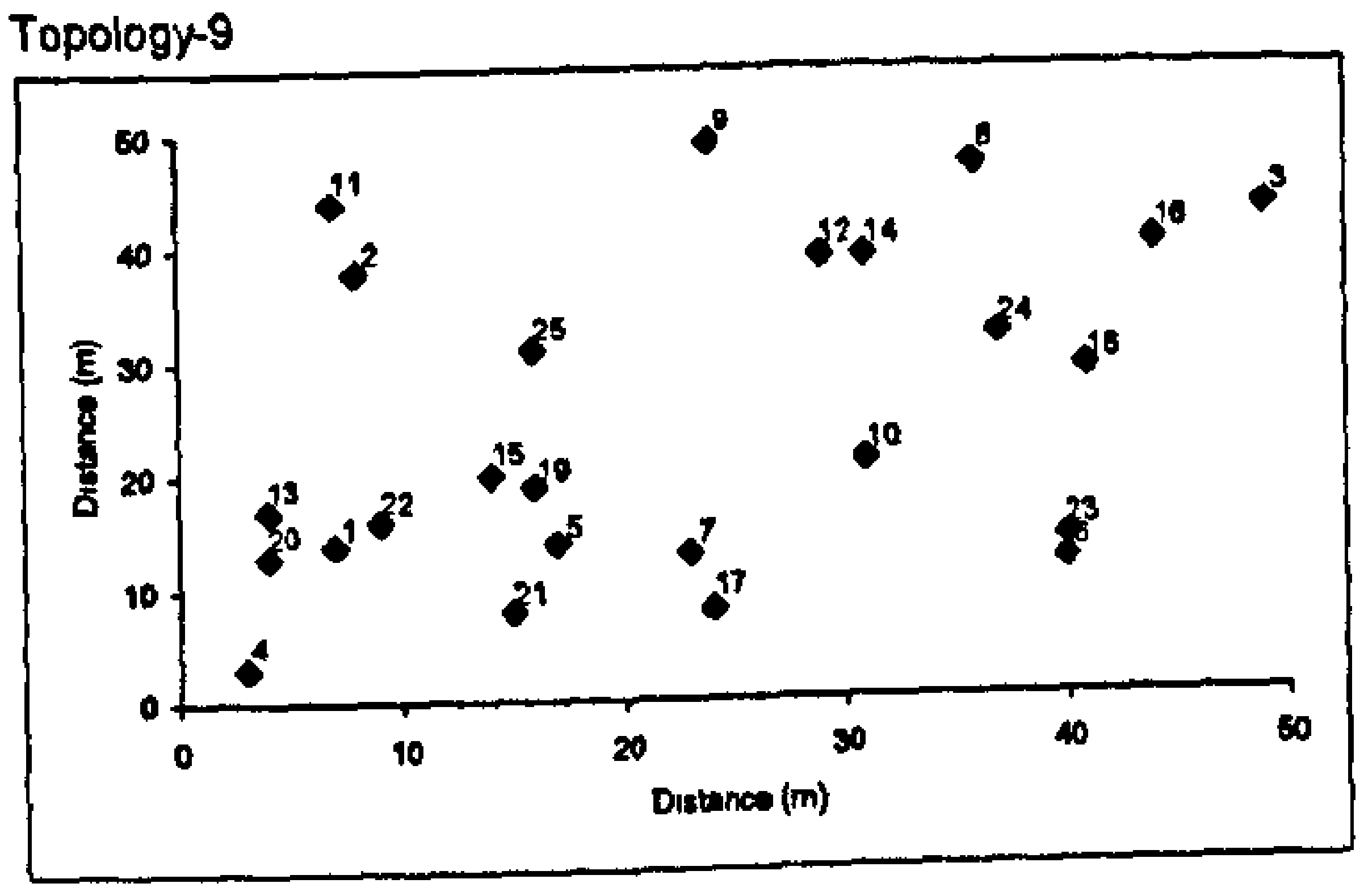
Topology-7



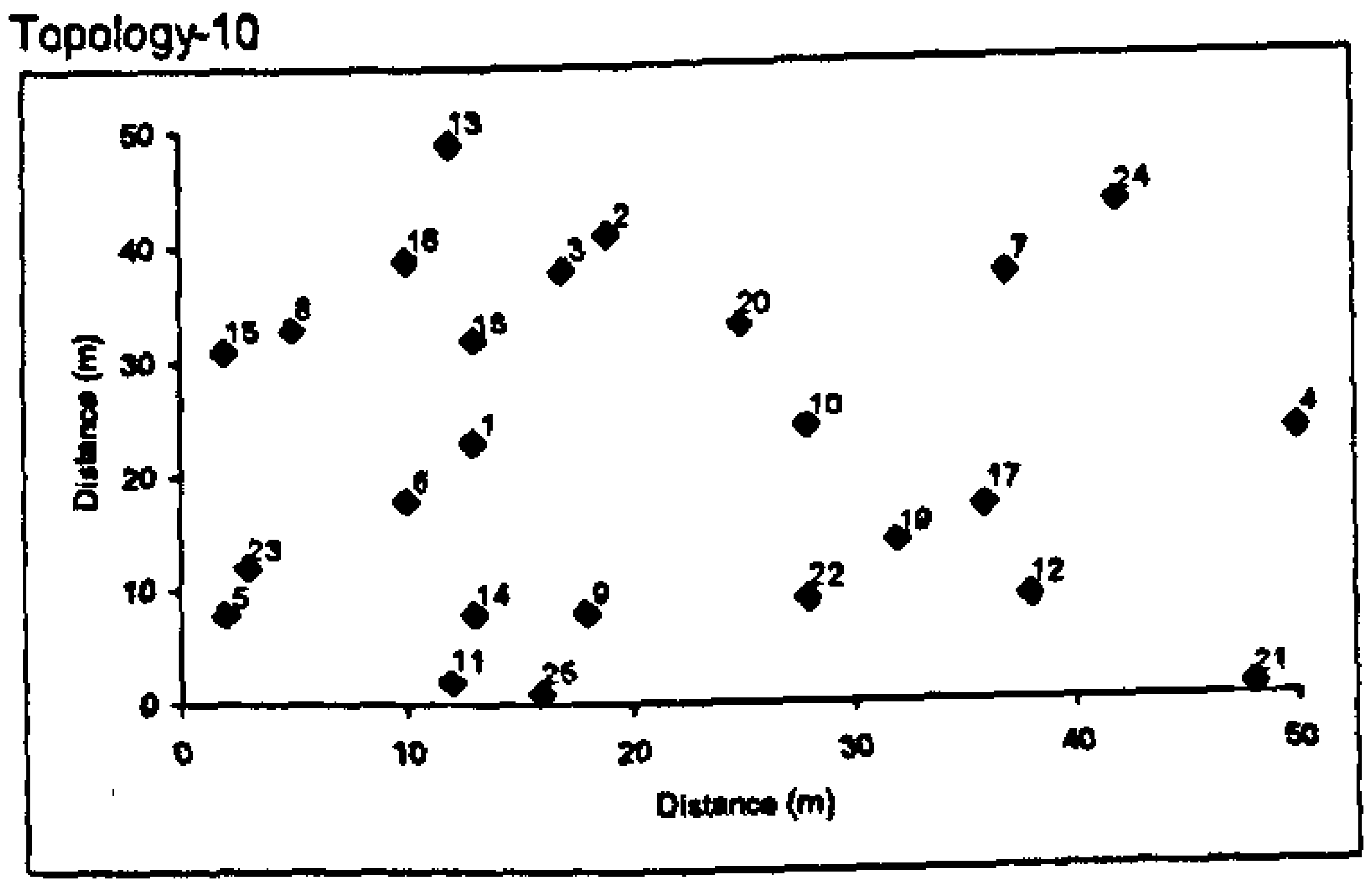
Topology-8



Topology-9



Topology-10



Network lifetime in different network topologies (scenario A)

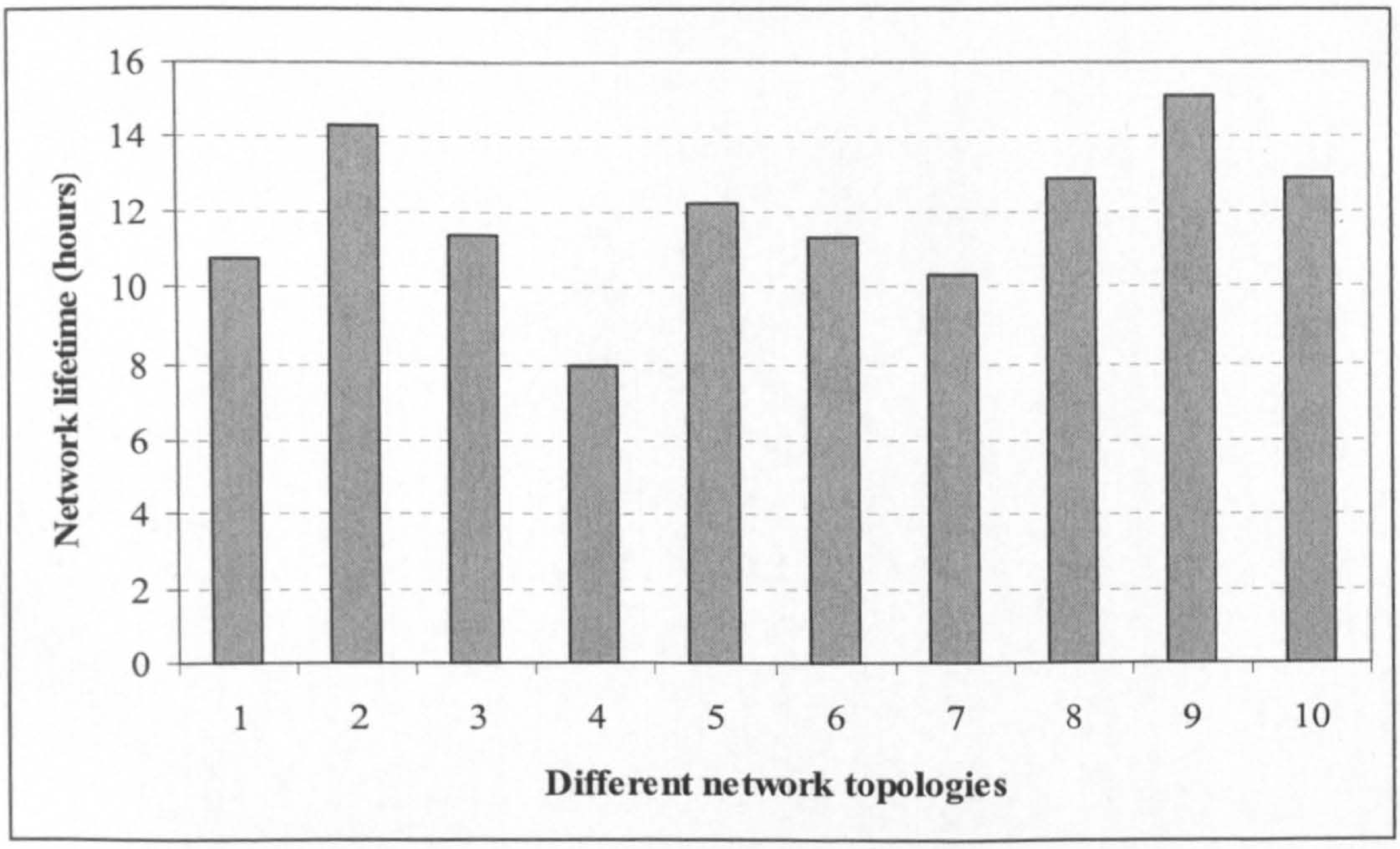


Figure N.1 Network lifetime in MPR

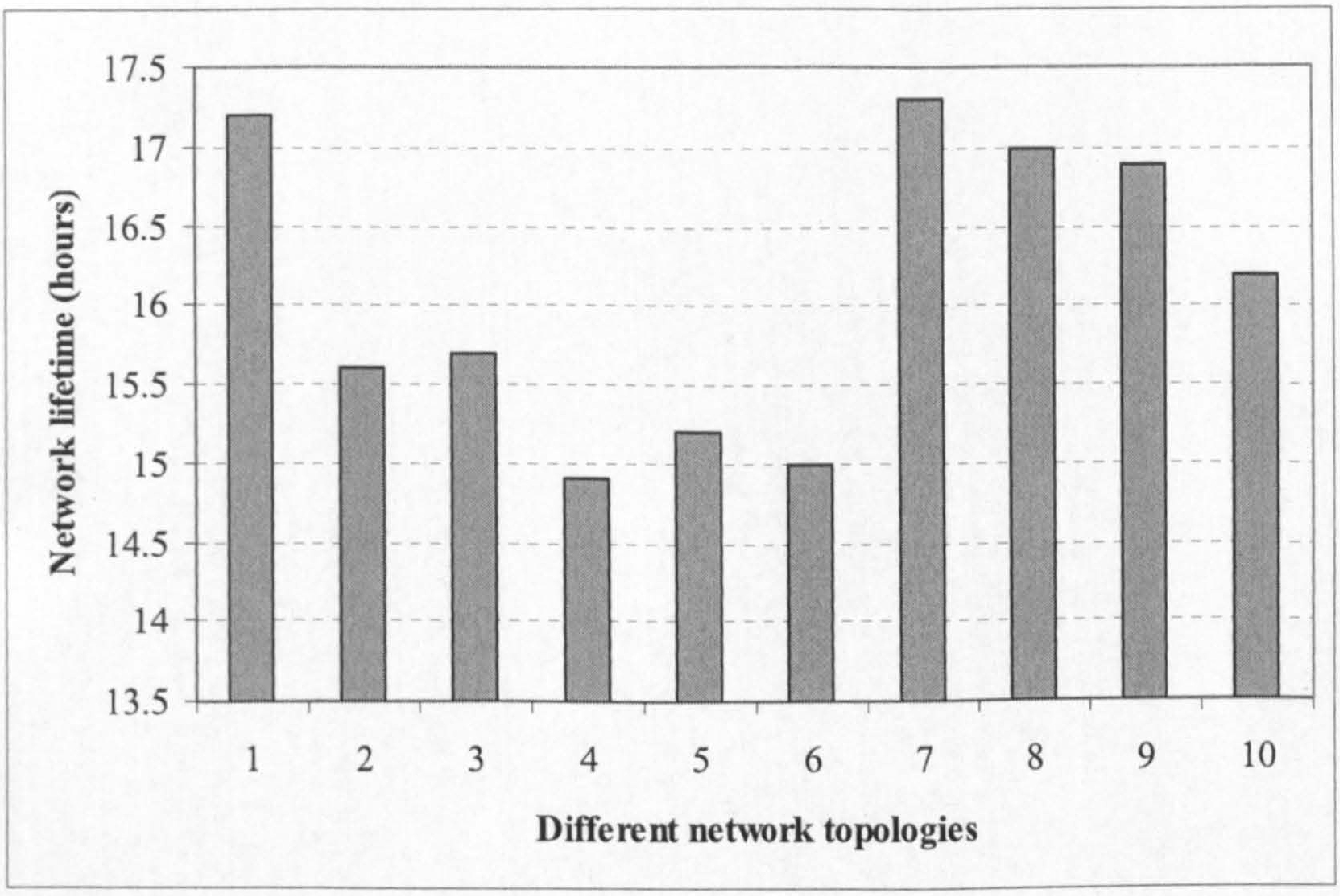


Figure N.2 Network lifetime in battery charge threshold (30%)

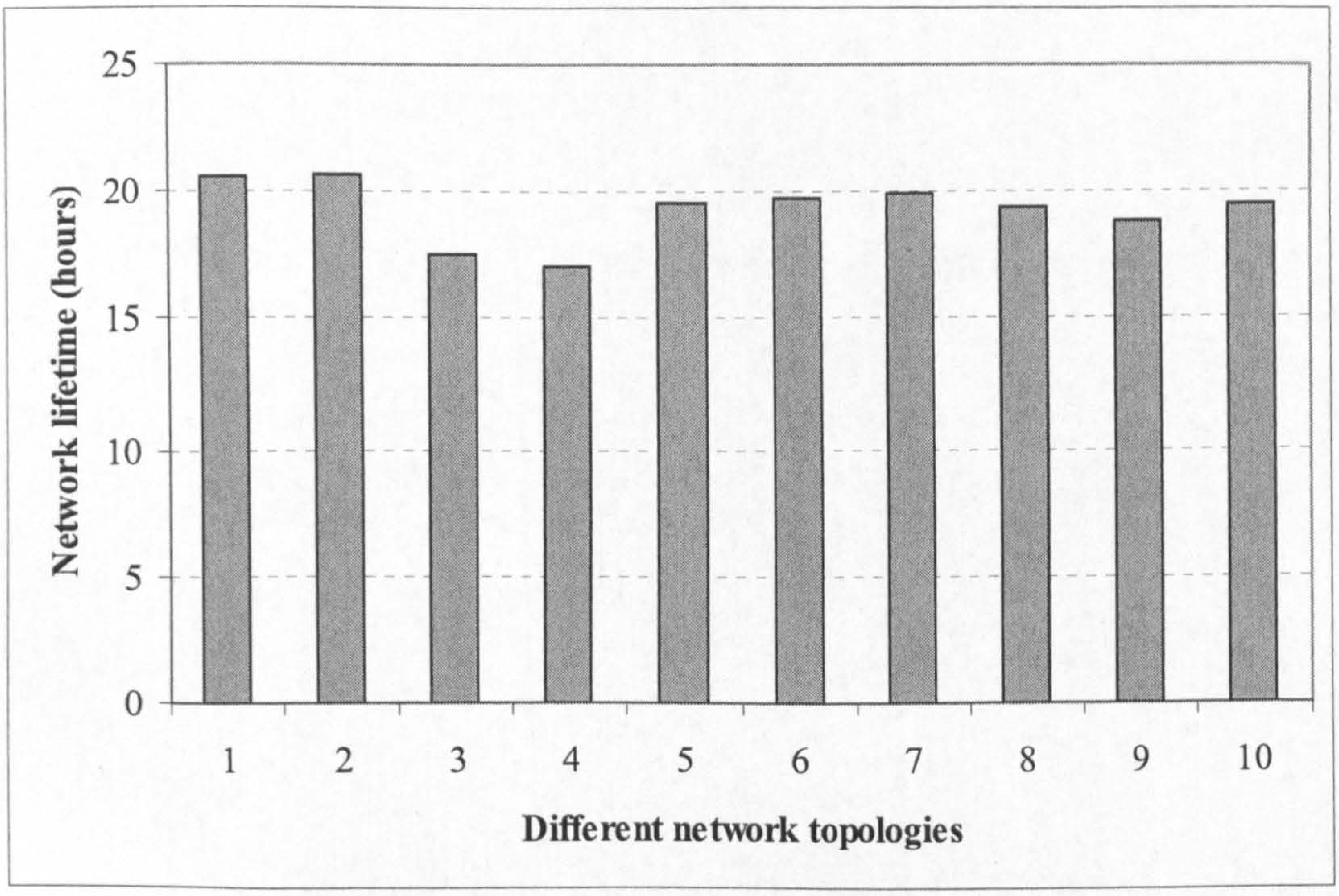


Figure N.3 Network lifetime in residual battery charge scheme

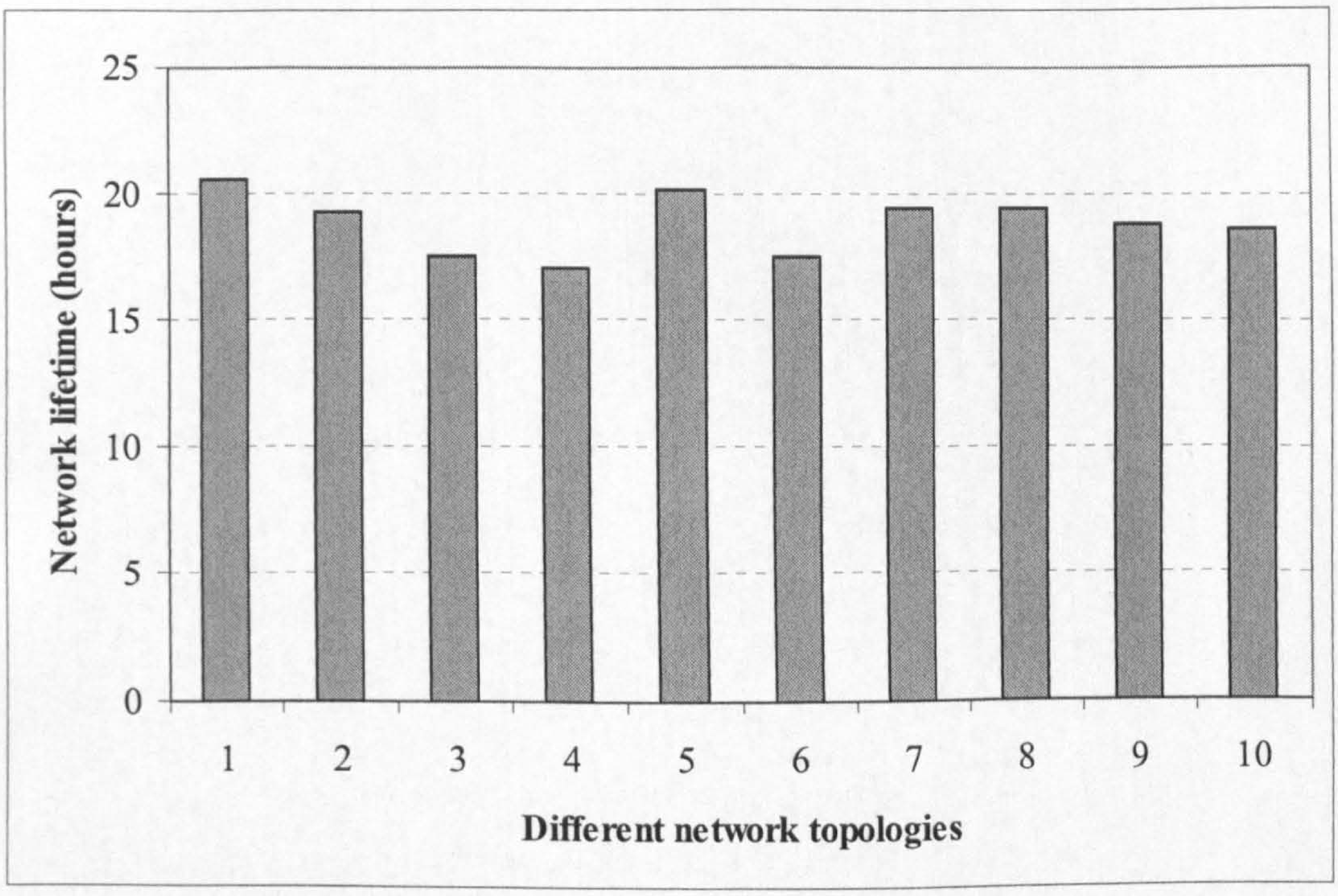


Figure N.4 Network lifetime in MPR/MBL
(inverse relationship, cross over point 600mAhrs, a=5)

Routing schemes	Avg network lifetime (Hrs.) (without throughput maintained)
Minimum power	12
Minimum power with a battery charge threshold (30%)	16
Residual battery charge	19
Power aware routing	18
MPR/MBL (exponential, cross-over point 600mAhrs, a=5)	19

Table N.1 Average network lifetime of various topologies