**Department of Electrical and Computer Engineering**

# Mitigating Hidden Node Problem in an IEEE 802.16 Failure Resilient Multi-hop Wireless Backhaul

**Pey San Nancy Chai**

**This thesis is presented for the Degree of**

**Doctor of Philosophy**

**of**

**Curtin University**

**May 2011**

# DECLARATION

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for award of any other degree or diploma in any university.

Signature: ……

Date: ……..

# ABSTRACT

Backhaul networks are used to interconnect access points and further connect them to gateway nodes which are located in regional or metropolitan centres. Conventionally, these backhaul networks are established using metallic cables, optical fibres, microwave or satellite links. With the proliferation of wireless technologies, multi-hop wireless backhaul networks emerge as a potential cost effective and flexible solution to provide extended coverage to areas where the deployment of wired backhaul is difficult or cost-prohibitive, such as the difficult to access and sparsely populated remote areas, which have little or no existing wired infrastructure.

Nevertheless, wireless backhaul networks are vulnerable to node or link failures. In order to ensure undisrupted traffic transmission even in the presence of failures, additional nodes and links are introduced to create alternative paths between each source and destination pair. Moreover, the deployment of such extra links and nodes requires careful planning to ensure that available network resources can be fully utilised, while still achieving the specified failure resilience with minimum infrastructure establishment cost.

The majority of the current research efforts focus on improving the failure resilience of wired backhaul networks but little is carried out on the wireless counterparts. Most of the existing studies on improving the failure resilience of wireless backhaul networks concern energy-constrained networks such as the wireless sensor and ad hoc networks. Moreover, they tend to focus on maintaining the connectivity of the networks during failure, but neglecting the network performance. As such, it calls for a better approach to design a wireless backhaul network, which can meet the specified failure resilience requirement with minimum network cost, while achieving the specified quality of service (QoS).

In this study, a failure resilient wireless backhaul topology, taking the form of a ladder network, is proposed to connect a remote community to a gateway node

located in a regional or metropolitan centre. This topology is designed with the use of a minimum number of nodes. Also, it provides at least one backup path between each node pair. With the exception of a few failure scenarios, the proposed ladder network can sustain multiple simultaneous link or node failures. Furthermore, it allows traffic to traverse a minimum number of additional hops to arrive at the destination during failure conditions.

WiMax wireless technology, based on the IEEE 802.16 standard, is applied to the proposed ladder network of different hop counts. This wireless technology can operate in either point-to-multipoint single-hop mode or multi-hop mesh mode. For the latter, coordinated distributed scheduling involving a three-way handshake procedure is used for resource allocation. Computer simulations are used to extensively evaluate the performance of the ladder network. It is shown that the three-way handshake suffers from severe hidden node problem, which restrains nodes from data transmission for long period of time. As a result, data packets accumulate in the buffer queue of the affected nodes and these packets will be dropped when the buffer overflows. This in turn results in the degradation of the network throughput and increase of average transmission delay.

A new scheme called reverse notification (RN) is proposed to overcome the hidden node problem. With this new scheme, all the nodes will be informed of the minislots requested by their neighbours. This will prevent the nodes from making the same request and increase the chance for the nodes to obtain all their requested resources, and start transmitting data as soon as the handshake is completed. Computer simulations have verified that the use of this RN can significantly reduce the hidden terminal problem and thus increase network throughput, as well as reduce transmission delay.

In addition, two new schemes, namely request-resend and dynamic minislot allocation, are proposed to further mitigate the hidden node problem as it deteriorates during failure. The request-resend scheme is proposed to solve the hidden node problem when the RN message failed to arrive in time at the destined node to prevent it from sending a conflicting request. On the other hand, the dynamic minislot allocation scheme is proposed to allocate minislots to a given node according to the

amount of traffic that it is currently servicing. It is shown that these two schemes can greatly enhance the network performance under both normal and failure conditions.

The performance of the ladder network can be further improved by equipping each node with two transceivers to allow them to transmit concurrently on two different frequency channels. Moreover, a two-channel two-transceiver channel assignment (TTDCA) algorithm is proposed to allocate minislots to the nodes. When operating with this algorithm, a node uses only one of its two transceivers to transmit control messages during control subframe and both transceivers to transmit data packets during data subframe. Also, the frequency channels of the nodes are pre-assigned to more effectively overcome the hidden node problem. It is shown that the use of the TTDCA algorithm, in conjunction with the request-resend and RN schemes, is able to double the maximum achievable throughput of the ladder network, when compared to the single channel case. Also, the throughput remains constant regardless of the hop counts.

The TTDCA algorithm is further modified to make use of the second transceiver at each node to transmit control messages during control subframe. Such an approach is referred to as enhanced TTDCA (ETTDCA) algorithm. This algorithm is effective in reducing the duration needed to complete the three-way handshake without sacrificing network throughput. It is shown that the application of the ETTDCA algorithm in ladder networks of different hop counts has greatly reduced the transmission delay to a value which allows the proposed network to not only relay a large amount of data traffic but also delay-sensitive traffics. This suggests that the proposed ladder network is a cost effective solution, which can provide the necessary failure resilience and specified QoS, for delivering broadband multimedia services to the remote rural communities.

# ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Kah Seng Chung, for his patience, guidance, and advice throughout the course of this research.

I would also like to express my gratitude to my co-supervisor, Dr King Sun Chan, for his supervision, help, and support throughout the course of study.

Also, I would like to express my appreciation to Curtin University for funding my study for 42 months through the Curtin International Postgraduate Research Scholarships (CIPRS). I would also like to thank Prof Syed Islam for his kind tuition fee sponsorship through the Department of Electrical and Computer Engineering. This thesis would have been impossible without these greatly-appreciated educational funds.

In addition, I would like to thank all the members in Communication Technology Research Group (CTRG) for providing a friendly and supportive working environment. I would also like to acknowledge the help and support of our administrative and technical staffs.

Last but not least, I would also like to express my gratitude to my family and friends for their unceasing encouragement, patience, and love which keep me through all the difficult times of this research.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ABBREVIATIONS

| | |
|---|---|
| BSS | Basic service set |
| CBR | Constant bit rate |
| CDS | Collision domain set |
| CH | Channel |
| ETTDCA | Enhanced two-channel two-transceiver distributed channel assignment |
| GHz | Giga Hertz |
| ICT | Information and communication technology |
| ID | Identification |
| IE | Information element |
| IEEE | Institute of Electrical and Electronics Engineers |
| IPTV | Internet Protocol television |
| LAN | Local area network |
| MAC | Medium access control |
| Mbps | Megabit per second |
| MSH-DSCH | IEEE 802.16 mesh distributed schedule message |
| MSH-NCFG | IEEE 802.16 mesh network configuration message |
| MSH-NENT | IEEE 802.16 mesh network entry message |
| MTU | Maximum transmission unit |
| NCTUns | National Chiao Tung University network simulator |
| OFDM | Orthogonal frequency division multiplex |
| PMP | Point-to-multipoint |
| QAM | Quadrature amplitude modulation |
| QoS | Quality of service |
| RN | Reverse notification |
| TCP-IP | Transmission Control Protocol – Internet Protocol |
| TDD | Time division duplex |
| TDMA | Time division multiple access |
| TTDCA | Two-channel two transceiver distributed channel assignment |
| TW | Three-way |

| | |
|---|---|
| UDP | User Datagram Protocol |
| VoIP | Voice over Internet Protocol |
| WiMax | Worldwide Interoperability for Microwave Access |
| WLAN | Wireless local area network |

# LIST OF SYMBOLS AND NOTATIONS

| | |
|---|---|
| $\lvert \bullet \rvert$ | Absolute value operator |
| $R_D$ | Data generating rate |
| $T_S$ | Duration of an OFDM symbol |
| $T_F$ | Frame duration |
| $x$ | Holdoff exponent |
| $x_j$ | Holdoff exponent of node $j$ |
| $x_k$ | Holdoff exponent of node $k$ |
| $\lvert Dx \rvert_{\min}$ | Minimum number of minislots required by a given node |
| $m_b$ | Number of bits in a minislot |
| $m$ | Number of bits that can be transmitted in an OFDM symbol |
| $N_B$ | Number of bits transmitted in a frame |
| $E[S_j]$ | Number of control transmission opportunities lost by node $j$ due to contention |
| $E[S_k]$ | Number of control transmission opportunities lost by node $k$ due to contention |
| $E[S]$ | Number of control transmission opportunities lost due to contention |
| $N_i^{fl}$ | Number of flows for node $i$ |
| $N_k^{fl}$ | Number of flows for node $k$ |
| $N_F$ | Number of frames |
| $n$ | Number of frames that can be allocated to a node |
| $\lvert Dx \rvert$ | Number of minislots |
| $N_i^s$ | Number of minislots allocated to node $i$ |
| $\lvert Dx \rvert_{candidate}$ | Number of minislots of a node after taking consideration of fairness and spatial reuse |
| $N$ | Number of nodes |
| $N_k^{known}$ | Number of nodes with known schedules |

| | |
|---|---|
| $N_k^{unknown}$ | Number of nodes with unknown schedules |
| $nbr(j)$ | Number of nodes within two-hop neighbourhood of node $j$ |
| $k$ | Number of OFDM symbols in one minislot |
| $M$ | Number of transmissions within a two-hop neighbourhood |
| $o_h$ | Overhead |
| $R_{PHY}$ | Physical layer data rate |
| $r_s$ | Request size |
| $\lfloor \bullet \rfloor$ | Rounding down to the nearest integer |
| $\lceil \bullet \rceil$ | Rounding up to the nearest integer |
| $Tload_t$ | Theoretical maximum traffic load |
| $T_{handshake}$ | Time needed to complete a TW handshake |
| $N_{fr}^s$ | Total number of minislots in a frame |
| $\alpha$ | Tuning parameter used for estimating $R_D$ |
| $\beta$ | Tuning parameter used for estimating $T_{handshake}$ |

# CHAPTER 1

# INTRODUCTION

## 1.1 SCOPE OF THE THESIS

Although information communication technology (ICT) services are prevalent in densely populated and developed urban areas, many geographic regions, particularly the sparsely populated remote areas, still remain without access to these services. Telecommunication service providers are not keen to extend their services to cover these areas, which are often regarded as commercially unviable due to the much lower demand for services compared to the urban areas [1]. Also, the provision of ICT services to the remote areas incurs high costs, as it requires an appropriate backhaul network for delivering broadband telecommunication services to remote communities from gateway nodes located in regional or metropolitan centres. Moreover, the deployment of wired ICT systems in remote areas is often challenging, due to the unavailability of existing infrastructure and the difficult topographical conditions, like mountainous terrain, valleys, swamp, etc.

Satellite communication has traditionally been used to provide telecommunication services to remote rural areas. Although it is able to provide flexible coverage, its total throughput is relatively low of the order of a few Mbps [2]. Often, its use incurs a high on-going subscription cost. Because of the large uplink and down-link distances, a satellite communication link tends to suffer from excessive round trip propagation delay which could be as long as half a second [3]. For this reason, echo cancellation is usually applied in satellite links to maintain acceptable quality for voice communication. On the other hand, it is desirable to investigate other alternative wireless backhaul technologies which could provide cost effective ICT services to remote rural regions. One possible solution is to make use of a terrestrial broadband wireless backhaul, which could be based on an off-the-shelf radio technology, such as the IEEE 802.16 standard. It is envisaged that such a wireless

backhaul would be able to support throughput in the order of tens of Mbps in conjunction with average delay of less than 150 ms with proper scheduling and routing [4].

Nevertheless, the wireless IEEE 802.16 backhaul network is susceptible to occasional failures due to local geographical and climate conditions, as well as interference caused by other extraneous electrical signals. In order to maintain undisrupted traffic transmission, even in the presence of network failures, it is essential to incorporate failure resilience in the design of a wireless backhaul network. Failure resilience is the ability of a network to perform its designated set of functions with minimal sacrifice of quality of service (QoS) when failures occur [5, 6]. This requires the network to be able to detect possible failures and redirect traffic away from failed network segment via alternative paths by adopting some form of traffic rerouting scheme. In addition to ensuring the successful arrival of traffic at the destination, it is also important to examine the effect of adding extra nodes and links to the network for providing failure resilience on the performance of the network during both normal and failure conditions.

Most of the existing research efforts tend to focus on improving the failure resilience of wired backhaul networks [7-13], but less investigate the wireless counterpart, although it has started to draw attention recently [14]. In fact, the design of failure resilient wireless backhauls, which requires the consideration of several constraints specific to the broadcast nature of wireless, such as interference and transmission power, is conspicuously more challenging compared to its wired counterpart. For this reason, the majority of existing studies tend to focus on achieving one of the various factors, for example, maintaining connectivity among the nodes during failure, but do not consider the possible increase of co-channel interference when additional nodes and links are incorporated for improving the failure resilience of a network. Also, most of these studies are devoted to looking at ways of improving the failure resilience of energy-constrained wireless sensor and ad hoc networks [14-20]. However, the operational requirements of these networks are quite different from wireless backhaul networks. As such, it calls for a better way to design a failure resilient wireless backhaul network to achieve a specified QoS requirement while

still meeting the specified failure resilience with minimum infrastructure establishment cost.

## 1.2 OBJECTIVES AND ORIGINAL CONTRIBUTIONS

The primary objectives of this research are:

a) To design a multi-hop wireless backhaul network that can tolerate multiple simultaneous node and link failures for delivering broadband telecommunication services to a remote community from a regional or metropolitan centre.

b) To investigate ways of realising a multi-hop wireless backhaul through the use of minimum number of intermediate nodes while meeting the necessary level of failure resilience.

c) To study the performance of the IEEE 802.16 coordinated distributed scheduling operating in a failure resilient multi-hop wireless backhaul network, and improve on any shortcomings, in order to achieve greater throughput and lower transmission delay during normal operation as well as under failure conditions.

In order to achieve these aims, the following novel contributions are the highlights of this research:

i) A failure resilient multi-hop wireless backhaul network, which has a high degree of resilience and can provide at least a backup path for every node pair, is proposed and presented in Section 4.3. Such a network is able to sustain multiple simultaneous node and link failures. Also, traffic in the network needs only to traverse a minimum number of hops when it is rerouted during failures. Moreover, the design of such network is realised with the use of a minimum number of nodes.

ii) It is identified that the IEEE 802.16 coordinated distributed scheduling suffers from severe hidden node problem in Section 4.4.3. A novel reverse

iii) A request-resend scheme is proposed to further reduce the hidden node problem, especially in the presence of a node or link failure. With this scheme, which is described in Section 5.3.1, a node will send a new request when it fails to get its requested minislots due to the hidden node problem. The request-resend scheme incorporated with the RN scheme is shown to be able to significantly reduce the hidden node problem in the proposed ladder network during normal condition. As a result, the throughput achieved is very close to the theoretical maximum traffic load that can be supported by the network.

iv) A dynamic minislot allocation scheme is formulated to dynamically allocate minislots to links according to their level of congestion during failure conditions. This scheme involves assigning a larger number of minislots to links that are required to reroute traffic. This is done by allocating a few non-consecutive minislot blocks to these links if continuous minislots are not available. The dynamic minislot allocation scheme, when operating in conjunction with request-resend and RN, is shown to be able to improve the throughput of the ladder network, independent of the number of hop counts, during failure condition. A detailed description of this scheme is presented in Section 5.3.2.

v) To further improve the throughput performance of the wireless backhaul network, a two-channel two-transceiver distributed channel assignment (TTDCA) algorithm is proposed to allocate minislots to the individual nodes equipped with two transceivers. With this algorithm, which is described in Section 6.2, all nodes will tune one of their transceivers to a common channel

vi) When operating with the TTDCA algorithm, a node only uses one of its two transceivers during the control subframe. An enhanced TTDCA (ETTDCA) algorithm is proposed in Section 6.4 to make use of the second transceiver to transmit grant information elements (IEs) in the control subframe. It is shown that this algorithm is effective in reducing the transmission delay without sacrificing the network throughput.

## 1.3 STRUCTURE OF THE THESIS

In Chapter 2, the type of failures commonly encountered in a wireless backhaul is discussed. This is followed by a description of conventional backhaul topologies and their shortcomings. Then, the essential factors that influence the design of a failure resilient wireless backhaul are examined before the current research efforts are reviewed. The design of wired and wireless failure resilient networks and the differences between the two are also identified.

Next, in Chapter 3, the physical and medium access control (MAC) layer specifications of the IEEE 802.16 standard are presented. This includes the description of the IEEE 802.16 mesh mode frame structure, coordinated distributed

scheduling, and the three-way handshake procedure. This is followed by an extensive literature review of existing techniques, which have been proposed for improving the performance of the coordinated distributed scheduling and three-way handshake, such as the alteration of holdoff exponent and holdoff base values, enhancement of data scheduling, as well as using multiple frequency channels for transmissions.

In Chapter 4, the various factors considered in the design of a failure resilient wireless backhaul network are examined. Then, a simple failure resilient ladder network is described in Section 4.3. The IEEE 802.16 three-way (TW) handshake is then applied for coordinated distributed scheduling in the proposed ladder network and the performance of the network is evaluated. It is shown that the TW handshake procedure suffers from a severe hidden node problem and thus a new reverse notification (RN) scheme is proposed in Section 4.4.4 to overcome the problem. Computer simulated results for the performance of the RN scheme during normal condition are presented in Section 4.4.5.

The performance of the RN scheme in the presence of a link or node failure is examined in Chapter 5. Two new schemes, namely request-resend and dynamic mini-slot allocation, are proposed for incorporation in the standard IEEE 802.16 coordinated distributed scheduling in conjunction with RN for use in the ladder network. The performance of the new schemes is then evaluated during both normal and failure conditions. Moreover, the effect of varying the buffer size and packet size on the performance of the ladder network is also examined. Furthermore, the performance of the ladder network is compared with a network which consists of two parallel paths of the same number of hop counts. This has been carried out with the networks operating under unidirectional and also bidirectional traffic transmissions.

In Chapter 6, a two-channel two-transceiver distributed channel assignment (TTDCA) algorithm is presented in Section 6.2. Then, an enhanced TTDCA (ETTDCA) algorithm is proposed in Section 6.4 to make use of both transceivers in a node to transmit control messages during the control subframe. The performance of such an algorithm during normal and failure conditions is evaluated in Section 6.5. The performance of the ladder network in the presence of bidirectional traffics is also

investigated. Finally, the major findings are reviewed and recommendations for future studies are made in Chapter 7.

# CHAPTER 2

# REVIEW OF FAILURE RESILIENT NETWORKS

## 2.1 INTRODUCTION

Backhaul networks are used to interconnect intermediate nodes or base stations to gateway nodes which are located in regional or metropolitan centres. Often, these backhaul networks link several distant communities in urban, suburban, or rural areas, with the aim to provide users in these areas with broadband and telecommunication services. These networks are designed to carry large amounts of data and real time traffics such as voice and video, which flow to or from the gateways via the intermediate base stations. An example of a backhaul network is illustrated in Figure 2-1.

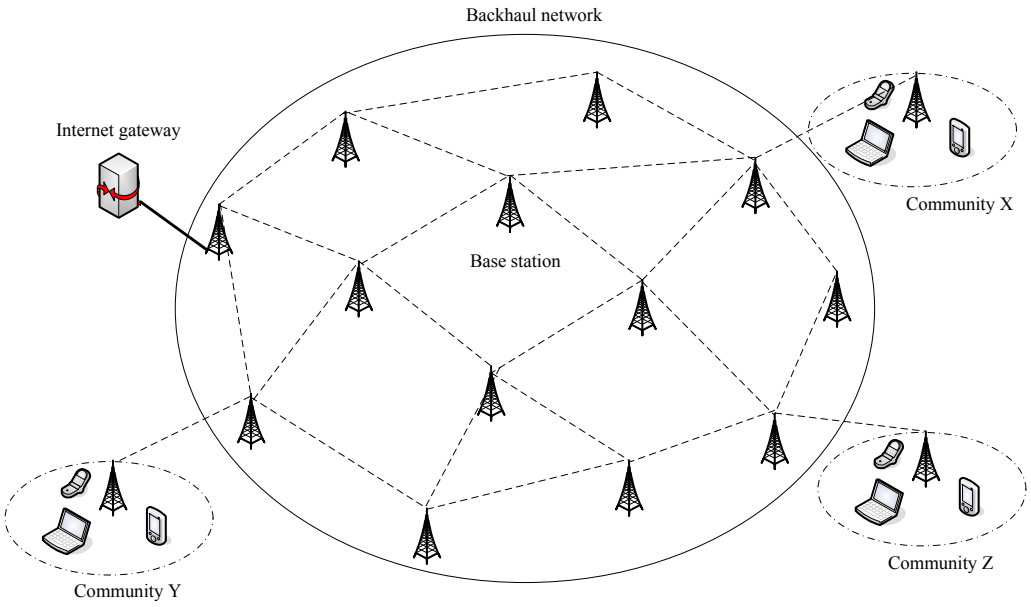



Figure 2-1 An example of backhaul network.

Traditionally, backhaul networks are established using metallic cables, optical fibres, microwave or satellite links. Both metallic cables and optical fibres are guided transmission media where signals are directed to travel within the physical limits of

the medium. In metallic cables, signals are transported in the form of electrical current whereas in optical fibres signals are transmitted in the form of light. These transmission media are reliable and secure as they are less affected by interference, fading, and weather. Moreover, they can provide high transmission speed, making them the most appealing options to be deployed in highly populated urban and suburban locations. Occasionally, microwave or satellite links are used in areas where wired connections are unavailable or difficult to install because of terrain constraints. Microwave transmissions can be carried out in licensed or unlicensed frequency bands. Often, backhaul networks are established in licensed frequency bands to provide reliable and good quality of service (QoS). On the other hand, satellite links are established between earth-orbiting communications platforms and base stations on earth. Satellite links could be readily set up to provide flexible coverage. However, high establishment costs and long propagation delays tend to make satellite communication a less attractive option for use in backhaul networks [2]. For example, typical propagation delay for satellite links is around 270 ms plus processing delay, and this is more than the generally acceptable end-to-end delay of 150ms for real time services, such as voice telephony [21].

Today, fast growing bandwidth demands for telecommunication services, particularly high speed data services such as internet access, video calls and even mobile television, are putting great strains on many existing communication backhaul networks. Also, there is a tendency nowadays to narrow the so called digital divide between the highly populated urban centres, and often remote and small rural communities. These have put pressure on service providers to upgrade the capacities of backhaul networks in high population density urban areas, and extend the coverage range in order to offer broadband services to remote population centres. As such, service providers are now looking beyond traditional backhaul technologies for a potentially low cost and easy to deploy solution, such as the use of wireless technologies, to meet the challenge of providing cost effective broadband services to the remote rural communities.

The use of wireless technologies makes rapid deployment of low cost backhaul networks possible in locations which are difficult to reach, or with low population density. Furthermore, a wireless backhaul can be readily expanded by introducing

additional base stations to the existing network, if required, to increase its coverage and accommodate more users.

The open broadcast nature of wireless communications tends to make operation of a wireless backhaul network susceptible to local geographical and climatic conditions, as well as interference caused by other extraneous electrical signals. The former can adversely alter the channel characteristics of a wireless link by giving rise to excessive waveform distortion and attenuation. On the other hand, interference can cause degradation in the signal-to-noise ratio. As a result, a wireless link may lose its connectivity leading to what is normally termed link failure. Furthermore, a loss in network connectivity can also occur in the event of faulty radio equipment at a node. This is usually referred to as a node failure. In order to maintain undisrupted traffic transmission, even in the presence of network failures, it is crucial to incorporate failure resilience in the design of wireless backhaul networks [22, 23]. Failure resilience is defined as the ability of a backhaul network to perform its designated set of functions when failures occur, even though this might mean a momentary degradation in QoS [5, 6]. As such, a failure resilient wireless backhaul network should be able to detect possible failures and minimise potential traffic losses, by redirecting traffic away from the failed network segment through the use of some form of traffic rerouting scheme. Traffic rerouting is only feasible if alternative paths are available between the source and sink nodes in the network. Hence, in order to design a failure resilient wireless backhaul network, it is necessary to consider the types of possible failures and other factors, such as the network cost, level of connectivity, and rerouting schemes.

## 2.2 BACKHAUL FAILURES

Two types of transmission failure can occur in a backhaul network, namely node and link failures. A communication link failure is detected when traffic fails to arrive at a particular node within a specified period of time. Such a failure may occur due to congestion caused by heavy traffic, node mobility, the hidden station problem, or damage to an upstream node [24]. Although a link failure may involve a single or multiple links, in practice, single-link failure is more likely to happen [25, 26]. On the other hand, a node failure is caused by equipment breakdown, which in turn

could be associated with one or more factors, such as power failure, natural disaster, incorrect maintenance or human errors. The occurrence of a node failure is generally less frequent than for a link failure. However, a node failure always gives rise to more than one link failure. For this reason, greater effort is normally devoted to minimise node failure, and provide necessary backup in case when this happens [26, 27].

Backhaul failures can further be categorised as either permanent or transient failures. The former will usually involve physical repair to restore the normal network operation. Node failures may in most cases be considered as permanent failures. Conversely, transient failure, which is common in link failures, can often be recovered automatically in a very short period of time. As such, its impact on the network performance is often transitional. In fact, there is little attempt to try to correct for transient failures, as such a move may induce undesirable oscillatory behaviour or instability to the network [26].

Occasionally, the occurrence of a single failure at a link or node may lead to multiple such failures in the network. An example is when traffic directed to a failed node is redistributed to its neighbouring nodes with the possible consequence of causing some of these nodes to become overloaded with traffic and failed. Since the occurrences of these events are interrelated, the resulting failures are referred to as correlated. Other causes of correlated failures are natural disasters, terror attacks, and power outages.

## 2.3 CONVENTIONAL BACKHAUL TOPOLOGIES

Backhaul networks are usually established using one or a combination of the basic topologies, such as chain, tree, star, and ring, as shown in Figure 2-2(a) ─ (d), respectively.

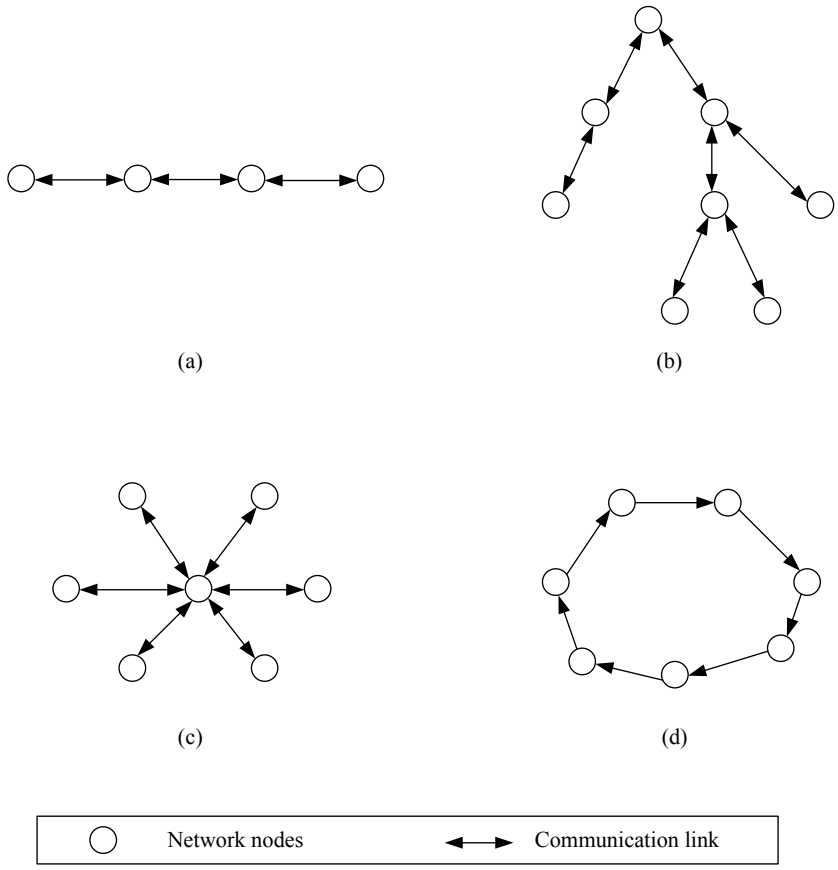



Figure 2-2 Basic network topologies: (a) chain, (b) tree, (c) star, and (d) ring.

These basic topologies only provide a single link between a given pair of nodes, thus making them susceptible to failure in traffic delivery. For a network to be able to sustain link or node failure, alternative paths between each pair of nodes will have to be established using additional links and nodes. For example, in Figure 2-3(a), a bypass route is provided for rerouting traffic in case the original link between the source and destination nodes fails. On the other hand, when the intermediate node between the source and destination nodes, as shown in Figure 2-3(b), fails, traffic can then be rerouted via an alternative route, which bypasses the failed node, to arrive at the destination. However, the introduction of any extra link and node will incur additional cost to network deployment. For this reason, the design of a failure resilient backhaul network calls for careful planning to ensure that the specified level of failure resilience is able to be achieved with minimum infrastructure establishment cost.

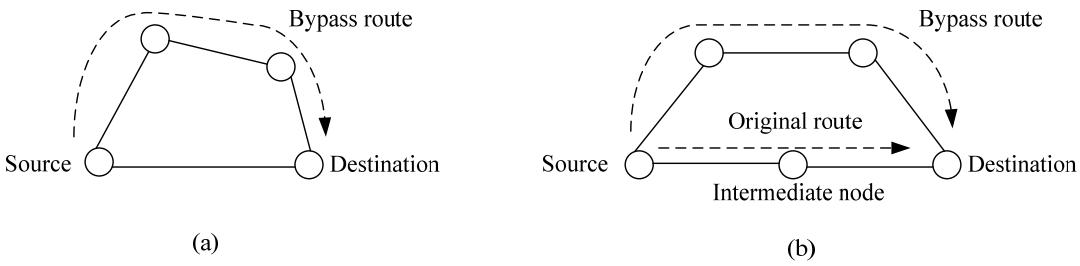



Figure 2-3 Addition of a bypass route for: (a) a link failure in the original route; and (b) a node failure in the intermediate node.

## 2.4 CONSIDERATIONS IN THE DESIGN OF FAILURE RESILIENT NETWORK

The design of a failure resilient backhaul network often begins with the use of a physical topology graph to represent the network. An example of a topology graph that constitutes various network nodes and links is shown in Figure 2-4. In this case, each node on the graph represents a network element, such as a base station. Also, possible communication links between network nodes are indicated by the lines or arcs connecting individual pairs of nodes. When designing a failure resilient backhaul network, it is necessary to consider the following factors:

- Deployment cost
- Type of failures
- Level of connectivity or failure resilience
- Traffic rerouting strategy
- Required quality of service

These various factors will now be further elaborated in the following sections.

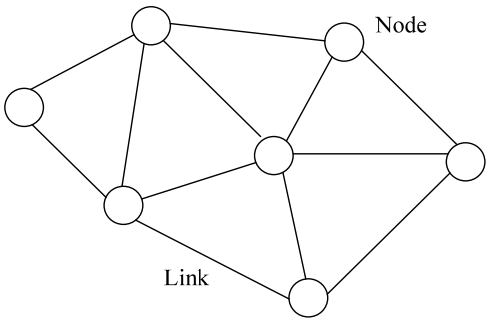



Figure 2-4 An example of a physical topology graph for a network.

### 2.4.1 Network Deployment Cost

The deployment cost of a network is generally associated with the total costs of setting up all the necessary network nodes and links. They include costs of equipping a node with the required hardware, and the materials used to implement a physical link, such as metallic or optical cable in the case of a wired network. In addition, there are labour costs involved in installing these nodes and links. The latter depends very much on the geographical terrain where the actual installations take place. Both of these material and labour costs are highly variable, depending on numerous factors, such as types and vendors of equipment, and geographical sites of individual nodes and links. For this reason, generalisation of the network deployment budget is not practicable. Moreover, it is reasonable to assume that the deployment cost is likely to grow with the number of nodes and links required. In order to keep the deployment cost low, it then becomes necessary to design a network which is capable of meeting the specified performance with as few nodes and links as possible [7] .

### 2.4.2 Type of Failures

Topology design of a failure resilient backhaul network is also influenced by the type of failures encountered. In the case of a node failure, backup paths that are typically node-disjointed or not sharing their origin with a failed node, are provided for rerouting traffic [28]. For example, as illustrated in Figure 2-5(a), Route 2, which is node-disjointed from Route 1, is used to provide backup for the failure of node B or C. On the other hand, link disjoint paths are introduced for rerouting traffic during a

link failure. In this case, a link disjoint path is one which does not share any segment of the route involving the failed link. As shown in Figure 2-5(b), Route 4 is considered as link disjointed from Route 3, and it may be used as backup path for the two links on Route 3.

Often, a node disjoint path may also provide backup for links on the original route as well as the links adjoining the route. Referring again to Figure 2-5(a), Route 2, which is a node disjoint path, can also provide backup for the link from node A to node B, and that from node C to node D. Likewise, as shown in Figure 2-5(b), Route 4, being a link disjoint path, may also act as a backup route for the nodes along the original route, i.e., Route 3, say when node Q malfunctions.

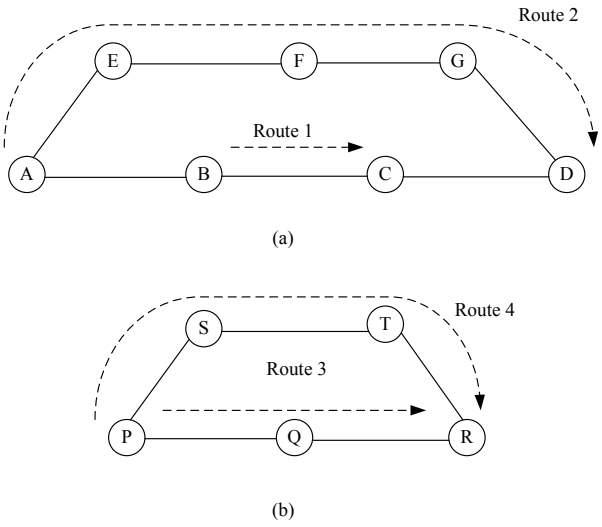



Figure 2-5 Backup paths: (a) Route 2 is node-disjointed from Route 1; and (b) link-disjointed routes, i.e., Route 3 and 4.

Furthermore, node or link disjoint paths are introduced as either dedicated or shared protection paths [29]. In the case of dedicated path protection, each individual node or link is protected with one or more dedicated alternative paths. In the event of a failure, traffic is rerouted via the alternative paths. Figure 2-6(a) shows an example in which two dedicated backup paths are provided to protect node Q, and the transmission link between nodes P and Q. In this way, a large amount of network resource will need to be devoted to provide individual backups for all the nodes and links in a backhaul network. Consequently, the practice of providing dedicated path

protection becomes very costly. An alternative approach is the shared protection scheme, which allows several nodes or links to make use of a small set of backup paths. In the simple example of Figure 2-6(b), a shared backup path is used to reroute traffic when there is a failure of either node Q or the link from node P to node Q. Although the use of shared protection paths often results in savings in network costs, it does not necessary guarantee the availability of sufficient restoration capacity in the event of multiple node and link failures. Again, referring to Figure 2-6(b), if both node Q and the link between node S and node T were to fail, traffic from node P will not be able to reach node R. This shortcoming may instead be overcome through the use of a combination of dedicated and shared protection as proposed in [30]. Such an approach is adopted in this study by providing a dedicated backup path to protect against a possible failure in a given network route. In addition, shared path protection is introduced to ensure an alternative route is available in case a subsequent failure also occurs in the dedicated backup path.

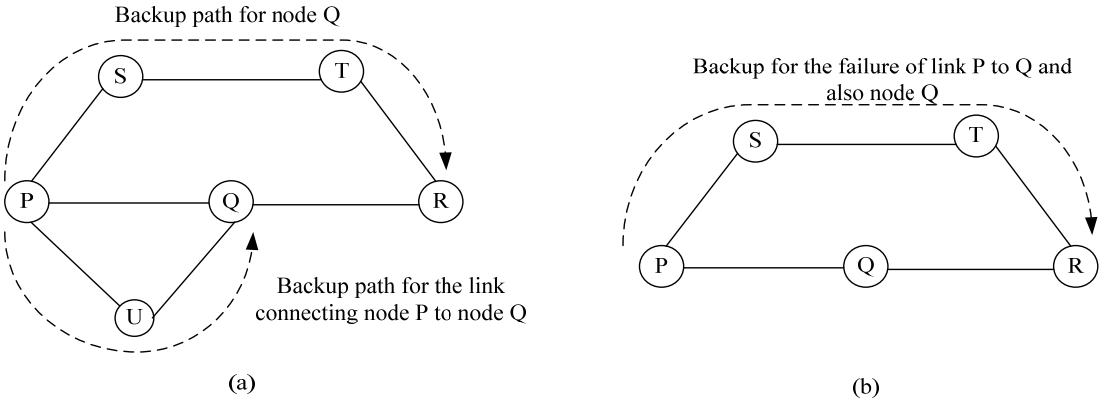



Figure 2-6 The backup paths used for rerouting traffic during the failure of node Q or the link between node P and Q can be provided as: (a) two distinct dedicated backup paths; or (b) a shared backup path.

### 2.4.3 Level of Connectivity

To enable a network to sustain multiple link and node failures, the level of connectivity among the network nodes has to be increased. Level of connectivity refers to the number of links incident on a node. For example, when a node is connected to at most two neighbouring nodes, it is said to have low level of

connectivity. With this arrangement, a given node will at least have a backup link for rerouting its traffic when one of its neighbouring nodes fails. However, it will be disconnected from the network if both of its neighbouring nodes were to fail simultaneously. This shows that a low connectivity network can only sustain a single node or link failure.

In order to cope with multiple simultaneous failures, a network node has to be connected to three or more neighbouring nodes. In other words, the network is said to have high level of connectivity. Because of the extra number of nodes and links needed to achieve the high level of connectivity, it is expected that the deployment of such a network will also incur a higher cost. In practice, there is often a trade off between connectivity and survivability requirements, and much research efforts must be devoted to arriving at an optimum balance between these two factors. For networks that are used to support critical applications, such as a communications backhaul, high connectivity becomes crucial to ensure these networks are able to remain functional even in the presence of multiple link and node failures.

### 2.4.4 Traffic Rerouting Strategy

When a node or link fails, traffic is rerouted via an alternative path, which has been established based on the adopted traffic rerouting strategy. One approach is to reroute traffic from the node immediately before the failed node or link, as shown in Figure 2-7(a). Such a strategy is referred as local rerouting. Alternatively, the upstream node may send a failure notification message back to the traffic source, as illustrated in Figure 2-7(b), which in turn will decide upon an appropriate alternative path for rerouting the traffic. Often, the choice of the alternative path is made based on its ability to support the required throughput and low delay [31]. Such an approach is commonly known as the path rerouting strategy, and is more capacity-efficient. However, its operation demands that the source node has complete information of the instantaneous traffic conditions within the network, and this may be too difficult to realise due to high complexity [11, 32, 33]. Furthermore, the failure notification message may need to propagate via multiple hops before it arrives at the source node. This may then introduce excessive delay which could prove to be unacceptable for delay sensitive applications. For this reason, local rerouting is often adopted in large

scale networks. This is especially true for a network which is prone to link failures, such as in a wireless backhaul [6].

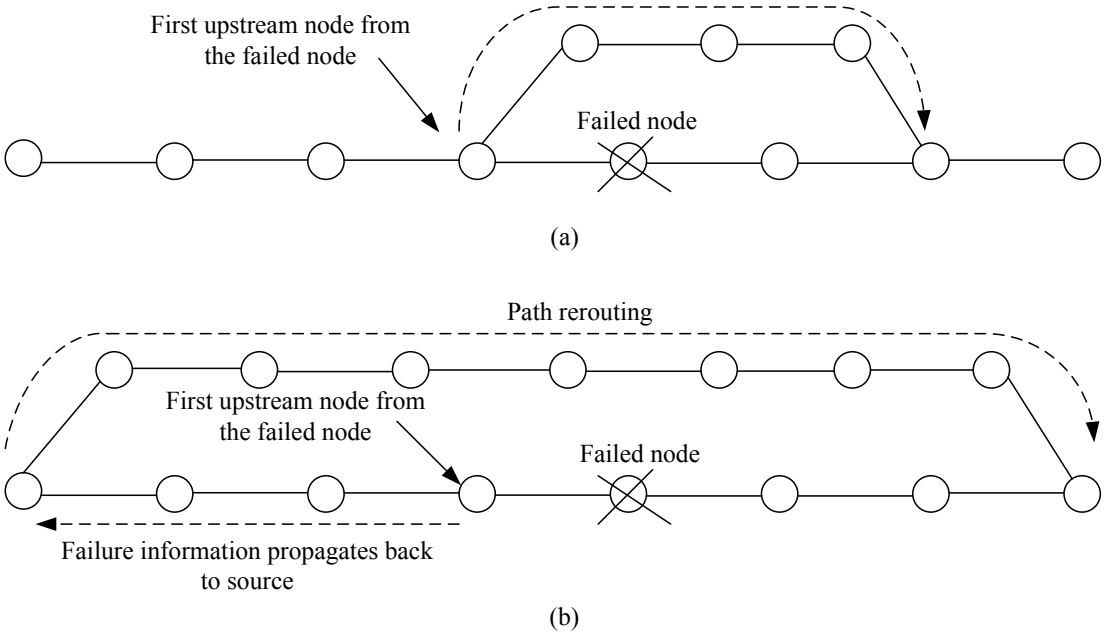



Figure 2-7 Two rerouting strategies: (a) local rerouting, which reroutes traffic from the node immediately before the failed node; (b) path rerouting that requires the transmission of failure information back to the source node. It will then select an appropriate rerouting path.

### 2.4.5 Quality of Service

Besides ensuring successful arrival of traffic at a destination node during a node or link failure, a network is required to maintain good QoS in terms of throughput and delay. This suggests that any backup path used will need to have sufficient capacity allocated to it for rerouting the additional traffic. Often, a dedicated backup path needs only to have sufficient capacity allocated for rerouting traffic when a specific node or link fails. On the other hand, the capacity of a given link or node along a shared backup path should be adequate for restoring traffic in the event of failures in a number of nodes or links. This may be done by either pre-reserving a certain capacity before a failure event or adaptively allocating the required capacity when a

failure occurs. The latter will experience a longer restoration time, but it can increase the capacity utilisation efficiency to achieve a higher network throughput [34].

Often, traffic needs to traverse a longer path when it is rerouted in the event of a network failure. As such, rerouted traffic is likely to incur a higher transmission delay. In a conventional circuit-switched backhaul, traffic is transmitted continuously via a dedicated transmission channel, and the transmission delay in most cases is assumed to be small and not considered as a design factor. On the other hand, when a more sophisticated channel access protocol, such as the IEEE 802.16 three-way handshake [35], is used to properly coordinate traffic transmissions traversing the various network nodes, each individual link is likely to encounter a longer connection set up time. Consequently, traffic transmission over a large number of hops will incur a high transmission delay [12]. For this reason, during network design, it is essential to consider ways of achieving transmission paths with a minimum number of hops.

In addition, a node in a wireless backhaul tends to suffer from a higher level of transmission interference, compared to its counterpart in a wired network, due to the broadcast nature of the wireless environment. Interference occurs when a given node lies within the transmission range of other nodes sharing the same frequency channel [36]. One way of minimising such interference is to have adjacent nodes transmit at different time instances. However, this means that each individual node will only have access to a small amount of transmission time. This situation becomes even worse, when extra nodes and links are introduced to provide backup paths for failure resilience. Consequently, this leads to a reduction in the amount of traffic that can be transmitted by a given node within a certain period of time. Such a shortcoming may be overcome by having adjacent nodes to adopt a different carrier frequency. In this way, adjacent nodes are able to transmit simultaneously their individual traffic streams with minimal mutual interference.

## 2.5 EXISTING FAILURE RESILIENT TOPOLOGY DESIGN

A majority of backhaul networks nowadays are based on the use of metallic cables and optical fibres. As such, most of the published studies are still focusing on ways of improving failure resilience in wired backhaul networks. Moreover, the requirement of failure resilience has also been incorporated into other networks, such as access networks, cellular networks, sensor networks, etc [12, 13, 15-19, 28]. Some of the approaches adopted to achieve failure resilience in these networks will now be reviewed. Such approaches serve as useful guidelines in the design of failure resilient wireless backhauls.

### 2.5.1 Wired Networks

The design of failure resilient wired backhaul networks is often carried out in two stages [7-9]. As described in [7], the locations of individual nodes in a failure resilient network are first identified during the first phase of the design procedure. Figure 2-8(a) shows an example of these node locations. Next, these nodes are numbered in no particular order, as shown in Figure 2-8(b). Communication links are then gradually added to the network until each individual pair of nodes has attained the minimum specified number of node disjoint paths. Figure 2-8(c) shows the topology obtained from the first design phase with an assumed cost of 243. This topology is then optimised for cost in the second phase of the design procedure to produce the final outcome, as illustrated in Figure 2-8(d), which yields a lower cost of 242. The design process is said to be completed, once an optimised topology that meets the requirement of the specified minimum number of node disjoint paths is obtained. A latter study in [10] further extends the design procedure to also incorporate the specification of minimum number of link disjoint paths between each node pair in the network topology. Often, it is difficult to perform the optimization process in a two-phase design procedure even for a small scale network. In this case, heuristic algorithms, such as the approximation methods for the travelling salesman problem [37, 38] and the “Bootstrap” heuristic [39], have been proposed to simplify the design problem and arrive at an appropriate network solution.

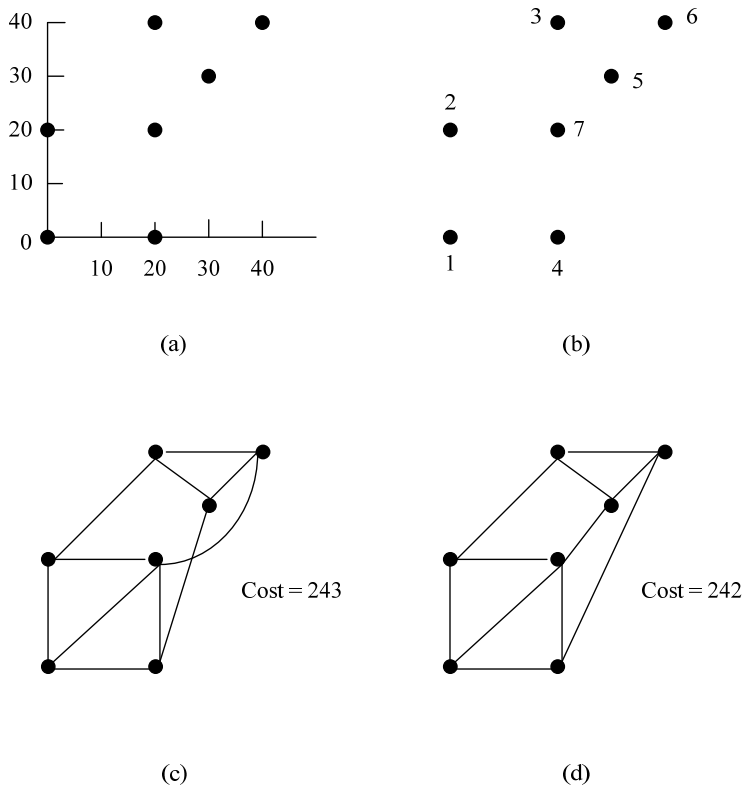



Figure 2-8 An example of a two-phase failure resilient wired network design: (a) locations of the nodes are specified; (b) random numbering of the nodes; (c) an example of the design solution after the first phase; (d) the final outcome after optimising the cost in the second phase.

A different approach to designing a failure resilient network is described in [11-13]. In the first instance, a minimum cost network is derived. This is then followed by a second phase which involves the introduction of additional backup paths to enhance the network reliability. Furthermore, the design procedure described in [12] also takes into consideration traffic demands and QoS requirements by introducing flow-balance constraints, and limiting the number of hops for each backup route in an attempt to reduce path delay and improve throughput. For large scale networks, computationally efficient heuristic algorithms have been proposed for speeding up the design process.

The use of two-phase design procedures enables the design of failure resilient networks based on mesh topology [11, 12]. On the other hand, a self-healing ring topology is proposed in [13]. Both the mesh and ring topologies are popular candidates for failure resilient networks, and they have also been adopted in several

other studies [5, 40-43]. In case of a mesh topology, nodes are geographically dispersed and each node is connected to two or more neighbouring nodes, as shown in Figure 2-9.

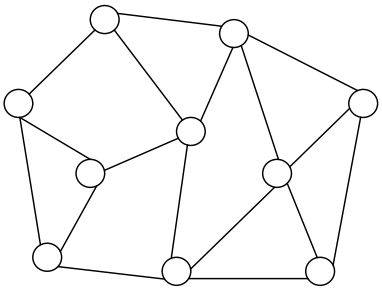


Figure 2-9 Mesh topology.

Mesh topology is particularly attractive for a coverage area where users are quite evenly distributed, such as in cellular networks. Although such a topology offers a very high level of connectivity to ensure network reliability, its adoption may not always be cost-effective. For example, in the case of a backhaul network, which is intended for connecting a remote community to a metropolitan centre, there are few or no users located along the areas covered by the intermediate nodes. In this situation, those intermediate nodes are used merely to relay traffic. Now, if a large number of intermediate nodes is deployed to form a mesh network for improving failure resilience, it will result in a large amount of network resources being wasted.

Unlike other network topologies, traffic travels only in one direction in a ring topology. A self-healing ring topology is constructed as a double-ring structure with the traffic flowing in opposite directions around the two rings. During normal operation, only one ring is used to transport traffic from one node to the next, as illustrated in Figure 2-10(a). However, when a link fails, traffic will be rerouted from the faulty link to travel in the opposite direction along the backup ring towards the destination node, as shown in Figure 2-10(b). In this case, the alternative path is much longer than the direct path. This suggests that traffic under failure condition will experience a longer delay to reach its destination. As such, a self-healing ring network is not suitable for traffic which is delay sensitive, such as voice traffic.

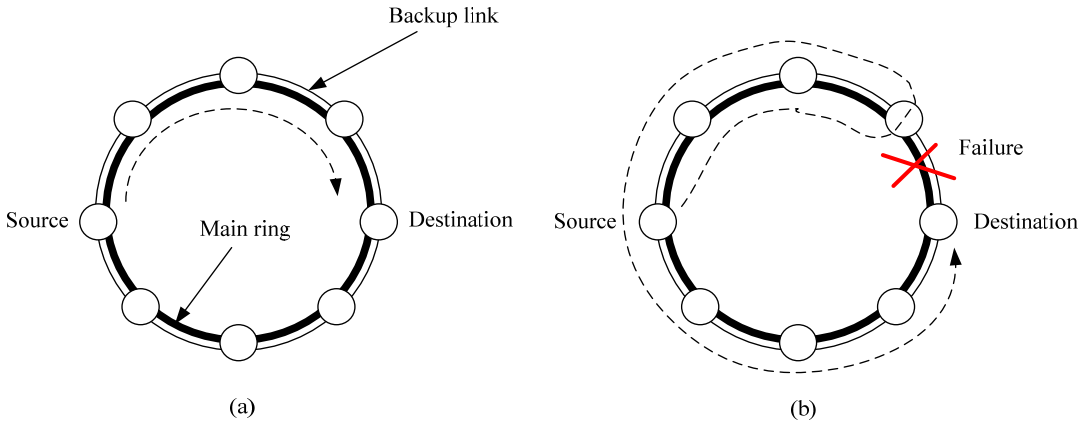



Figure 2-10 Traffic flow in a self-healing ring: (a) under normal operation; and (b) in the presence of a faulty link.

A tree topology has been proposed in [44] to overcome some of the shortcomings arising from transmission delay in a self-healing ring topology. An example of a tree network, equipped with parallel backup links to provide protection against possible link failure, is shown in Figure 2-11(a). It is shown in Figure 2-11(b) that during a link failure, traffic is more likely to make use of a shorter rerouting path to arrive at the destination node. However, it is also possible that some unprotected branches in the topology may fail, thus causing service disruption to certain segments of the network.

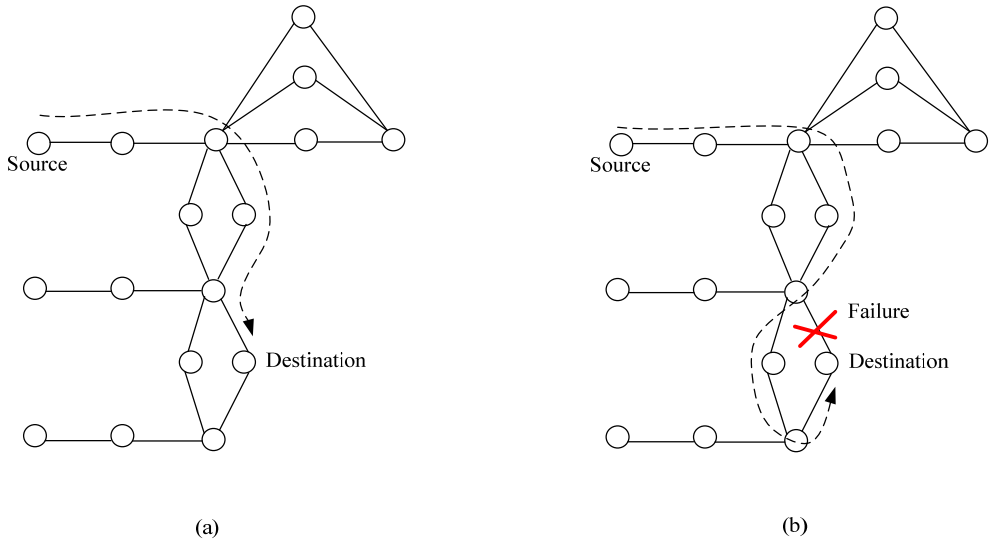



Figure 2-11 Traffic flow in a tree topology equipped with parallel backup links: (a) under normal operation; and (b) in the presence of a faulty link.

It is possible to design a failure resilient network by carrying out the optimisation process involving all the different considerations, such as cost, level of reliability, etc., in a single stage. For instance, in [45], with the specified geographical locations of the individual nodes and their traffic demand, a joint optimisation of topology design and bandwidth capacity allocation has been proposed for minimising the total network cost of a survivable optical mesh network. It is shown that a cost saving of 28 % can be achieved by adopting this approach for the design of a 10 node network.

Furthermore, it is proposed in [46] to jointly optimise both the network topology and spare capacity allocation. It also investigates the effect of using local and path rerouting strategies on the network cost. It is shown that the use of local rerouting will incur a slightly higher cost in link installation and capacity in the order of about 3 %. This slightly higher cost is considered acceptable as the implementation of local rerouting is much simpler than that for path rerouting [32].

All the studies referred to in the above paragraphs concern the design of failure resilient networks starting from scratch. Moreover, a study described in [47] looks at ways to augment an existing network with additional links, so as to increase its failure resilience. Instead of focusing on complete augmentation by providing at least two link-disjoint or node-disjoint paths for every node pair, as done in most other studies [48-50], it introduces backup paths to different parts of the network according to their required levels of protection. Also, it considers several other practical networking constraints during the augmentation process, including the influence of the geographical terrain on the cost of laying additional cables. In order to minimise the cost, shared backup paths are used. A similar study, which involves the augmentation of an existing radio access network to make it failure resilient, is presented in [51]. In this case, redundant links are introduced to the original network on an incremental basis, so as to create a partially meshed and restorable topology. In doing this, several practical issues, such as capacities of individual links, network cost, and traffic rerouting strategy, have also been considered.

### 2.5.2 Wireless Networks

Most current research effort on the design of failure resilient wireless network tends to be directed to sensor networks, ad hoc networks, and local area networks (LANs). With wireless sensor and ad hoc networks, individual nodes are often powered by batteries, and this constraint on available energy will influence the design of such networks. In order to save energy, wireless sensor and ad hoc networks are often established with their nodes located in clusters [15-17, 19, 52]. In each cluster, nodes are supposed to directly connect with each other, and one of these nodes is powered by a higher capacity battery to act as the cluster head, which is responsible for monitoring the operations of all the nodes within the cluster. Also, it has the responsibility to take the necessary action in restoring the operation of the cluster when a failure occurs. This then means that all the other nodes of a given cluster will not be able to communicate with nodes in another cluster in the event that its cluster head malfunctions. In an attempt to minimise such an event, periodic checks are usually carried out to monitor the status of individual cluster heads [52]. To minimise the effect of a cluster head failure, it is possible to arrange for those nodes under its supervision to reconnect themselves to operate with other adjacent cluster heads, based on the backup information created during the time of clustering. Another approach is to have every node connected with at least two neighbouring cluster heads [16]. In this case, if one cluster head fails, inter-cluster communication can still be maintained via the second cluster head. This strategy may be further extended to connect each node in a given cluster to a specified number of cluster heads [17].

Wireless local area networks (WLANs) are established in either ad hoc or infrastructure mode [53]. The former refers to a network of nodes, which can directly communicate with each other without going through an access point. Conversely, the infrastructure mode consists of several basic service sets (BSSs) where each BSS has one access point and several nodes. In this case, both intra-BSS and inter-BSS communications have to go through the access point. For the infrastructure mode, several studies have been devoted to searching for ways to restore connectivity between individual nodes and the responsible access point, when the latter happens to fail [54, 55]. One way is to reconnect the affected nodes back to the network by relaying their traffic through a bridging node, which is in communication with at

least one other access point [54]. However, such a scheme will only work when there are bridging nodes available within the transmission range of the affected nodes. Alternatively, the remaining access points can increase their transmission power to avoid any nodes from being disconnected from the network [55].

For a WLAN, which is operating in ad hoc mode and equipped with directional antennas, the study described in [56] proposed to improve the failure resilience of the network by ensuring each node is connected to two or more neighbouring nodes. In this case, the maximum number of neighbouring nodes of a node is bounded by its number of available directional antennas. An extension of this work is presented in [57], which examines how each node pair could be assigned the maximum possible capacity when the network is operating normally, or in the presence of network failures.

Furthermore, an iterative method is proposed in [58] for estimating the optimal number of backup nodes required to achieve a certain level of failure resilience in a given topology. With this scheme, each time after an additional backup node is introduced into the network, computation is carried out to estimate the degree of improvement in reliability that could be achieved. This is followed by a comparison with the additional cost incurred to achieve this improvement. As long as it is possible to derive more gain from the network reliability over the cost incurred, the process of adding an extra backup node will continue. Now, when the shortest path routing scheme is used, it is possible that not all the backup nodes will be involved in traffic transmission. In addition, this study does not consider the influence of adding backup nodes on network performance metrics, such as throughput and delay.

## 2.6 SUMMARY

Owing to their ease of deployment and low establishment cost, wireless backhauls have gained increasing popularity as an alternative option to provide modern broadband services to remote areas, which have little or no existing wired communication infrastructure. However, wireless backhaul networks are vulnerable to failures. These failures include node and link failures, which can happen in

isolation or simultaneously. It is possible for a single link or node failure to cause a large segment of a network to fail. This is particularly true when the network is established using either a chain, tree, star or ring topology. In order to ensure undisrupted operation even in the event of link and node failures, alternative paths are introduced to reroute traffic away from such failures. However, the need for extra links and nodes to establish alternative paths is likely to increase the network deployment cost. This suggests that careful consideration and planning are needed for designing a cost effective failure resilient wireless backhaul network.

This chapter considers several factors which can impact the design of a failure resilient wireless backhaul network. These factors include network cost, type of failures, level of connectivity, traffic rerouting strategies, and achievable QoS. As a wireless backhaul network is used to serve a large population of users, it must have high failure resilience and be able to provide acceptable QoS. For it to be cost effective, its deployment cost should also be kept to a minimum. However, some of these design factors tend to conflict with one another, and a sound compromise is often needed to arrive at a practical and cost effective solution. For example, the use of less nodes in a network is likely to lead to lower cost and better throughput coupled with lower delay. However, it may not allow the specified level of failure resilience to be achieved. Table 2-1 shows how the four design factors can be met through the quantity of nodes used in a network.

Table 2-1 The influence of the design factors on the number of nodes.

| **Design factor** | **Number of nodes used** |
|---|---|
| Network cost | Small |
| Capacity | Small |
| Interference | Small |
| Failure resilience | Large |

The use of a small number of nodes is likely to reduce the network cost. At the same time, each individual node will have a larger share of channel capacity, and it will also experience less interference. On the other hand, a high degree of network failure resilience will call for the use of a larger number of nodes. It remains a challenge to

find a way of arriving at an optimum number of nodes which will simultaneously satisfy all these design factors.

So far, the majority of the published papers tend to focus on improving failure resilience in wired networks. For those concerning failure resilient wireless networks, the focus is directed to sensor and ad hoc networks. Such studies are largely aimed at searching for ways to maintain network connectivity in the event of node failure, but do not consider the effects on achievable QoS due to the addition of links and nodes introduced for failure resilience. As wireless backhaul networks are likely to be used for connecting distant remote communities with regional centres, they would need to be highly reliable and efficient in the use of spectral resources. Also, they must provide good QoS to the users. Therefore, QoS must be included as an important factor in the design of a wireless backhaul network. This will be further examined in Chapter 4.

# CHAPTER 3

# IEEE 802.16 COORDINATED DISTRIBUTED SCHEDULING IN WIRELESS BACKHAUL

## 3.1 INTRODUCTION

Since wireless backhaul networks are deployed to serve a large population of users within an extended geographical area, a wireless communication protocol, which can provide high throughput over long distances, is required. There are two commonly used wireless communication protocols, namely IEEE 802.11 standard [59] and IEEE 802.16 standard [35]. The IEEE 802.11 standard, which was originally designed for indoor usage, can only support a coverage range in the order of a few hundred metres. As such, it is not suitable for use in long range wireless backhaul networks. On the other hand, the IEEE 802.16 standard [35], popularly known as Worldwide Interoperability for Microwave Access (WiMax), allows single-hop communications over a distance of up to 50 km [2]. Also, it supports multi-hop traffic transmissions to enable greater coverage range. Furthermore, the time division duplex (TDD) operating mode of the IEEE 802.16 standard works well with the inherently bursty and asymmetric data traffic of wireless backhaul networks [60]. With these desirable features, the IEEE 802.16 standard emerges as a promising wireless technology for use in wireless backhaul networks.

The IEEE 802.16 standard [35] specifies the air interface, including the physical layer and medium access control (MAC) layer specifications, for fixed and mobile broadband wireless access systems. The physical layer section of the IEEE 802.16 standard specifies several operating frequency bands. These frequency bands include the 10 – 66 GHz licensed band, as well as the license-exempt bands below 11 GHz [35]. In the case of operations at the license-exempt frequencies, the transmissions are likely to be subjected to excessive interference and co-existence issues. To ensure reliable operation, wireless backhaul networks are scheduled to operate in licensed frequency bands.

The IEEE 802.16 standard specifies two medium access modes, namely the point-to-multipoint (PMP), and mesh operating modes. The former is designed for communications between a base station and multiple subscriber nodes, which are one-hop away. As is common with radio communications, transmission from a base station to a subscriber is generally referred to as downlink transmission. On the other hand, an uplink transmission is associated with a signal being sent from a subscriber node to the base station. In the case of a downlink transmission, the signal is broadcast by the base station for reception by all the subscriber nodes located within its transmission range. Under this scenario, only the base station is transmitting and each subscriber station checks and retains only those data packets, which are addressed to it. However, in the uplink direction, it is possible that multiple subscriber stations may be competing for the same spectral resource by transmitting to the base station at about the same time. To achieve an orderly uplink transmission, special procedural rules are followed to coordinate transmissions from individual subscriber stations. In the case of PMP operation, the coordination is carried out centrally by the base station.

When operating in mesh mode, traffic from a gateway node often has to be relayed by other intermediate nodes before reaching its destination. In other words, mesh mode is normally associated with multi-hop traffic transmission, which is fundamental for any backhaul network. In addition, the physical layer of the mesh mode uses an orthogonal frequency division multiplex (OFDM) modulation scheme and the MAC layer is based on time division multiple access (TDMA). With the use of TDMA, time is partitioned into frames with duration of 2.5ms, 4ms, 5ms, 8ms, 10ms, 12.5ms, or 20ms, and each frame is further divided into time slots. An individual node may occupy one or more such time slots. In this way, multiple nodes sharing the same frequency channel will avoid transmitting at the same time. To coordinate transmissions by the nodes, two scheduling schemes are defined, namely centralised scheduling and distributed scheduling. The former requires a base station to gather information on resource requests from all mesh subscriber stations before allocating network resources to them. Although this can provide collision-free transmissions, the required messages from the subscriber stations would have to

propagate via multiple hops before arriving at the base station. Such a situation will inevitably cause a long connection set up time.

On the other hand, distributed scheduling may be used to coordinate data transmissions of the nodes in a fully distributed fashion, without requiring any interaction with the base station [61]. In addition, distributed scheduling can operate either uncoordinated or coordinated. Coordinated distributed scheduling allows a node to synchronise its transmissions with its two-hop neighbours in order to avoid collisions, thus making it more reliable than its uncoordinated counterpart. This makes coordinated distributed scheduling a better choice for use in wireless backhaul networks.

In this chapter, the frame structure adopted for mesh mode operation with the IEEE 802.16 standard is first reviewed in Section 3.2. Then, the way coordinated distributed scheduling operates is described in Section 3.3. The various factors, which will influence the performance of coordinated distributed scheduling, such as holdoff exponent, holdoff base, data scheduling, and multi-channel implementation, are discussed in Section 3.4. This is followed by a summary of the chapter in Section 3.5.

## 3.2 IEEE 802.16 MESH MODE FRAME STRUCTURE

Figure 3-1 shows the frame structure used for mesh mode operation with the IEEE 802.16 standard. A single mesh frame is made up of two segments, one for control messages and the other for data. The control subframe consists of MSH-CTRL-LEN transmission opportunities, each of which consists of seven OFDM symbols. On the other hand, the data subframe contains multiple minislots. A control subframe may be used for either network control or schedule control purposes. The network control subframe is used for transmitting mesh network configuration (MSH-NCFG) and mesh network entry (MSH-NENT) messages. These are used to facilitate network synchronisation and allow new nodes to join an existing network. For example, when a new node intends to join a mesh network, it must first attempt synchronisation with the network by listening to the MSH-NCFG messages to obtain the necessary network parameters. After that, it will send a MSH-NENT message to seek

permission to join the network. Moreover, mesh frames only carry a network control subframe on a periodic basis.

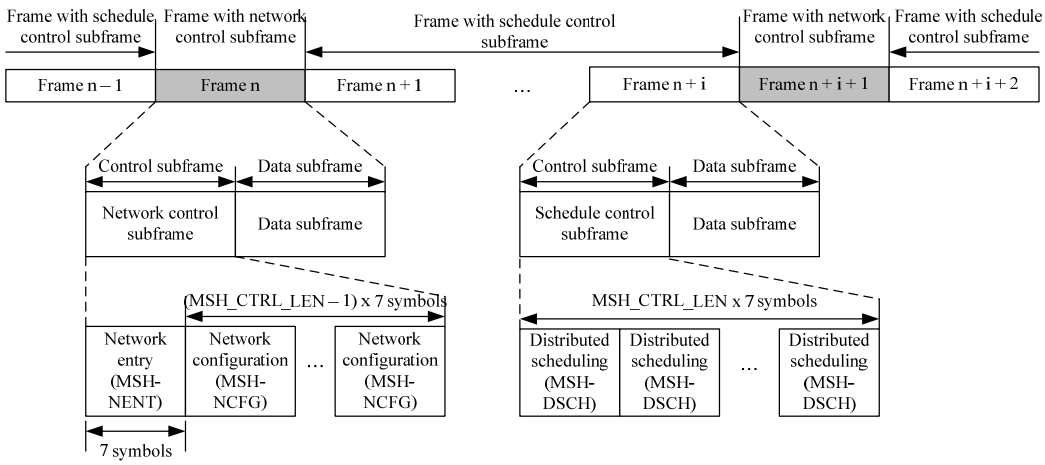



Figure 3-1 IEEE 802.16 mesh mode frame structure.

Next, a schedule control subframe is used to transmit mesh distributed schedule (MSH-DSCH) messages, which are responsible for coordinating transmissions of the control subframe as well as the data subframe. This will be further discussed in Section 3.3.

## 3.3 IEEE 802.16 COORDINATED DISTRIBUTED SCHEDULING

When operating in IEEE 802.16 coordinated distributed scheduling mode, each node of the mesh network will regularly transmit MSH-DSCH messages to broadcast not only its own transmission schedule for the next MSH-DSCH message but also those of its neighbours. A MSH-DSCH message carries two parameters, *Next Xmt Mx* and *Xmt Holdoff Exponent*. With these two parameters, the eligibility interval, *Next Xmt Time*, for a node to transmit its next MSH-DSCH message is determined, such that

$$2^{Xmt\ Holdoff\ Exponent} \times Next\ Xmt\ Mx < Next\ Xmt\ Time \le 2^{Xmt\ Holdoff\ Exponent} \times (Next\ Xmt\ Mx + 1) \quad \ldots(3.1)$$

where *Xmt Holdoff Exponent* can take on any integer value between 0 and 7 inclusive. From equation (3.1), the duration of the *Next Xmt Time* is $2^{Xmt\ Holdoff\ Exponent}$ transmission opportunities.

After a MSH-DSCH message is transmitted, the node has to holdoff for a period of *Xmt Holdoff Time*, which is equal to $2^{Xmt\ Holdoff\ Exponent+Holdoff\ Base}$ transmission opportunities. As specified in the IEEE 802.16 standard, the default value for *Holdoff Base* is 4. It then makes the first transmission opportunity after holdoff as its *temporary next transmission opportunity*. This also corresponds to the first opportunity which allows it to start competing for transmission resource within the control subframe. Meanwhile, its neighbours, located within two hops, may also compete for the same transmission opportunity under the following conditions:

1) These neighbouring nodes happen to have a *Next Xmt Time* value corresponding to the *temporary next transmission slot*.
2) The earliest possible transmission opportunity that these neighbouring nodes can transmit again after a holdoff period, also referred to as *Earliest Subsequent Xmt Time*, occurs at or before the *temporary next transmission slot* of the given node. Note that *Earliest Subsequent Xmt Time* is equal to *Next Xmt Time* + *Xmt Holdoff Time*.
3) The next transmission time for these neighbouring nodes being unknown.

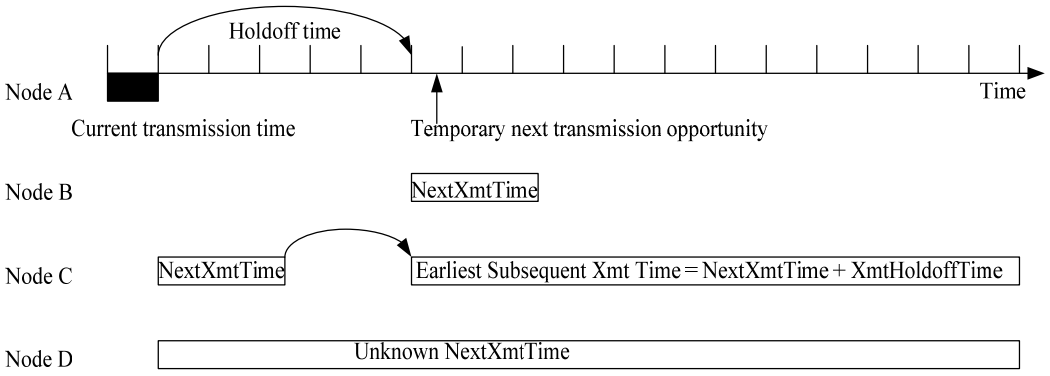



Figure 3-2 The IEEE 802.16 mesh control subframe contention.

For example, as shown in Figure 3-2, Node A is attempting to compete for the *temporary next transmission opportunity* with its two-hop neighbours, say Node B,

Node C, and Node D, which fulfil the above Condition 1, 2, and 3 respectively. Each of these four nodes will make use of the mesh election algorithm, as specified in the IEEE 802.16 standard, to generate a series of pseudo-random values, also known as the mixing values, based on the transmission opportunity number and the IDs of all the competing nodes. The node that produces the largest mixing value will then win the transmission opportunity by setting the *temporary next transmission opportunity* as its next transmission time slot. The winning node will also broadcast a MSH-DSCH message to inform its neighbours about the occupied transmission opportunity. Those nodes that lose out will repeat the same procedure in competing for the next available transmission opportunity, until they are able to finally win a transmission opportunity during the eligibility interval.

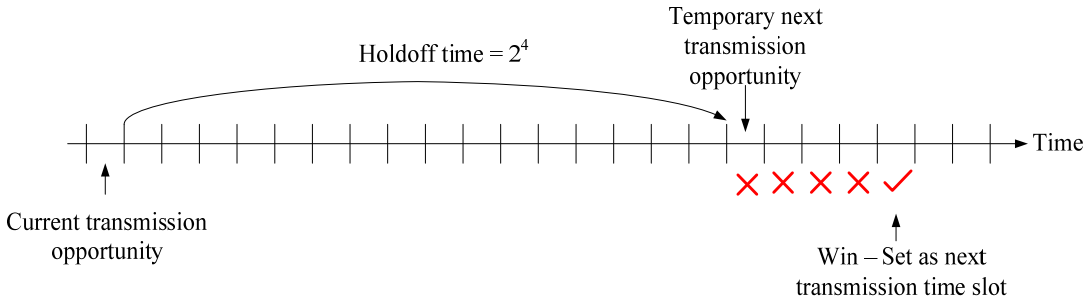



Figure 3-3 Control subframe hold off and contention.

Figure 3-3 illustrates that, after its current transmission, a given node will holdoff for sixteen transmission opportunities. It then sets the first transmission opportunity immediately after the holdoff period as its *temporary next transmission opportunity*. From the MSH-DSCH messages it received from its neighbours, it will be able to determine the number of competing nodes for the *temporary next transmission opportunity* as well as their respective *Next Xmt Time*. Subsequently, it makes use of the mesh election algorithm to determine whether it will be able to win the *temporary next transmission opportunity*. For the example as shown in Figure 3-3, after the given node fails to win the *temporary next transmission opportunity*, it will continue to keep on competing for the subsequent next transmission opportunities until it finally succeeds in obtaining one. It will then make this transmission opportunity the time slot for transmitting its control message. After that, it will broadcast a MSH-DSCH message to inform its neighbouring nodes about the

occupied transmission opportunity. By doing so, it is able to avoid collisions in MSH-DSCH message transmission with its two-hop neighbours.

Next, in order to ensure collision-free transmissions using the data subframe, the coordinated distributed scheduling, specified in the IEEE 802.16 standard, employs a three-way (TW) handshake procedure to set up connections between neighbouring nodes. This procedure is described in the following section.

### 3.3.1 Three-way (TW) handshake

The TW handshake process, as illustrated in Figure 3-4, makes use of the MSH-DSCH messages to relay the necessary scheduling information. The procedure begins when a node, say Node A, is ready to transmit its data packets. It first sends a request information element (IE), and the corresponding availability IE, to the intended receiving node, say Node B, using a MSH-DSCH message contained in the control subframe.

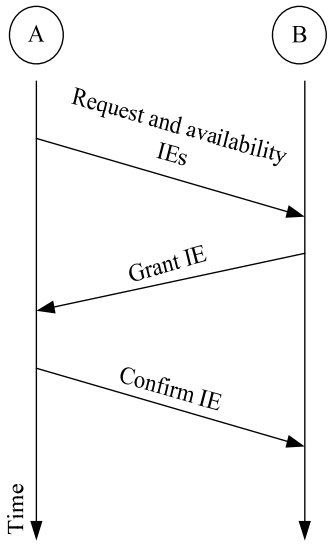



Figure 3-4 Three way handshake procedure in IEEE 802.16 coordinated distributed scheduling.

A MSH-DSCH request IE carries three parameters, namely Link ID, Demand Level, and Demand Persistence. The syntax of a MSH-DSCH request IE is shown in Table 3-1. Here, the Link ID is the ID assigned by the transmitting node to the link, along

which the request for bandwidth resource is made. The number of minislots requested for data transmission is specified by the Demand Level, while Demand Persistence indicates the number of frames in which the minislots requested is available. The values allowed for Demand Persistence are as tabulated in Table 3-2.

Table 3-1 MSH-DSCH request IE.

| **Syntax** | **Size** | **Notes** |
|---|---|---|
| MSH-DSCH_Request_IE() { | | |
| Link ID | 8 bits | |
| Demand Level | 8 bits | |
| Demand Persistence | 3 bits | |
| *reserved* | 1 bit | Shall be set to 0 |
| } | | |

Table 3-2 Demand persistence values.

| Demand persistence value | Indication |
|---|---|
| 0 | Cancel reservation |
| 1 | Single frame |
| 2 | 2 frames |
| 3 | 4 frames |
| 4 | 8 frames |
| 5 | 32 frames |
| 6 | 128 frames |
| 7 | Good until cancelled or reduced |

A MSH-DSCH availability IE, on the other hand, contains a list of consecutive minislots that are available for transmission by Node A. The syntax of a MSH-DSCH availability IE is tabulated in Table 3-3. It conveys six parameters, namely the Start Frame Number, Minislot Start, Minislot Range, Direction, Persistence and Channel. The Start Frame Number corresponds to the eight least significant bits of the number of the frame where the available minislots are found. The first available minislot with

the frame is specified by the parameter, Minislot Start, which also occupies 8 bits. Then, the Minislot Range indicates the number of minislots that are available to be granted. The indication of whether these minislots are otherwise unavailable, or free for transmission only, or free for reception only, or available for both transmission and reception, is conveyed by the value of the parameter, Direction, as given in Table 3-4. The 3 bits allocated for the parameter, Persistence, represent the eight allowed number of frames where available minislots are found. The values are the same as those tabulated in Table 3-2 for the Demand Persistence parameter used in the request IE. Finally, the parameter, Channel, signifies which of the 16 physical channels is adopted for the transmission.

Table 3-3 MSH-DSCH availability IE.

| **Syntax** | **Size** | **Notes** |
|---|---|---|
| MSH-DSCH_Availability_IE() { | | |
| Start Frame Number | 8 bits | 8 least significant bits of frame number |
| Minislot Start | 8 bits | |
| Minislot Range | 7 bits | |
| Direction | 2 bits | |
| Persistence | 3 bits | |
| Channel | 4 bits | |
| } | | |

Table 3-4 Direction values for MSH-DSCH Availability IE.

| Direction value | Indication |
|---|---|
| 0 | Minislot range is unavailable |
| 1 | Available for transmission in this minislot range |
| 2 | Available for reception in this minislot range |
| 3 | Available for either transmission or reception |

Upon receiving these two IEs from Node A, Node B will then determine whether the requested number of minislots is actually available for data reception. In the event that Node B decides that it is not possible to satisfy the resource request by Node A, it will either ignore the request, or allocate fewer minislots than those requested. Following this, Node B will send a MSH-DSCH grant IE back to Node A specifying the number of minislots to be allocated to it. The MSH-DSCH grant IE, as shown in Table 3-5, takes the same syntax as the MSH-DSCH request IE and MSH-DSCH availability IE with the exception that the parameter, Direction, now takes on the value of one, which indicates that it is a grant IE.

Table 3-5 MSH-DSCH grant IE.

| **Syntax** | **Size** | **Notes** |
|---|---|---|
| MSH-DSCH_Grants _IE() { | | |
| Link ID | 8 bits | |
| Start Frame Number | 8 bits | 8 least significant bits of start frame number |
| Minislot Start | 8 bits | |
| Minislot Range | 8 bits | |
| Direction | 1 bit | |
| Persistence | 3 bits | |
| Channel | 4 bits | |
| } | | |

Finally, the TW handshake process is said to have been completed once the requesting node, Node A, accepts the minislots allocation by sending a MSH-DSCH confirm IE back to Node B. The MSH-DSCH confirm IE is just a duplicate copy of the MSH-DSCH grant IE with the exception that the parameter, Direction, now takes on the value of zero to indicate that it is a confirm IE. While this TW handshake is going on between Node A and Node B, their neighbouring nodes are also able to overhear these exchanges of IEs. This allows them to pick up information on the status of Node A and Node B, particularly with regard to those minislots which have already been occupied. As a result, those neighbouring nodes will suspend their transmissions using the occupied minislots. In this way, the TW handshake

procedure is able to prevent transmissions of the neighbouring nodes from colliding with one another.

## 3.4 FACTORS INFLUENCING THE PERFORMANCE OF IEEE 802.16 COORDINATED DISTRIBUTED SCHEDULING

When operating in coordinated distributed scheduling mode, a node can only transmit data after successfully completing a TW handshake. It is, therefore, essential for a node to minimise the time it takes to complete the TW handshake if it is to avoid excessive delay in transmitting its data. This also suggests that the node and its respective receiving node must be able to exchange their IEs as soon as possible. This can only be achieved if both of these nodes adopt a short holdoff time, as well as being able to avoid losing out too many times in the contention for control transmission opportunities.

As mentioned in Section 3.3, the holdoff time of a node is given by

$$\text{Holdoff time} = 2^{Xmt\ Holdoff\ Exponent + Holdoff\ Base} \tag{3.2}$$

where the holdoff exponent can take on any integer value between 0 and 7 inclusive, and the default value for the holdoff base, as specified in the IEEE 802.16 standard, is 4. On the other hand, the number of control transmission opportunities lost due to contention, $E[S]$, is calculated based on the type of network topology, which can be correlated or general [61]. The former refers to a network of nodes, which are one-hop neighbours of each other. This differs from nodes in a general topology where they tend to have different numbers of neighbouring nodes. There are also those nodes, which do not have known schedules, and they are considered to be in competition for every transmission opportunity. For the correlated and general topologies, the $E[S]$ of a given node is calculated according to the following:

(i) For a correlated topology with nodes having the same holdoff exponent:

$$E[S] = (N-1)\frac{2^{x} + E[S]}{2^{x+4} + E[S]} + 1 \tag{3.3}$$

(ii) For nodes having a different holdoff exponent in a correlated topology:

$$E[S_k] = \sum_{j=1, j\neq k, x_j \geq x_k}^{N} \frac{2^{x_j} + E[S_k]}{2^{x_j+4} + E[S_j]} + \left( \sum_{j=1, j\neq k, x_j < x_k}^{N} 1 \right) + 1, \qquad k = 1,\ldots,N \tag{3.4}$$

(iii) For a general topology:

$$E[S_k] = \sum_{j=1, j\neq k, x_j \geq x_k}^{N_k^{known}} \frac{2^{x_j} + E[S_k]}{2^{x_j+4} + E[S_j]} + \left( \sum_{j=1, j\neq k, x_j < x_k}^{N_k^{known}} 1 \right) + N_k^{unknown} + 1, \quad k = 1,\ldots,N \tag{3.5}$$

where $N$ is the number of nodes in the network, $N_k^{known}$ is the number of nodes with known schedules, $N_k^{unknown}$ is the number of nodes with unknown schedules, $x$ is the holdoff exponent, and $x_j$ and $x_k$ are the holdoff exponents of node $j$, and node $k$, respectively. From equations (3.2) to (3.5), it is observed that both the holdoff time and $E[S]$ are influenced by the actual holdoff exponent value used. Furthermore, the holdoff time is also affected by the holdoff base value. As such, it becomes necessary to determine suitable values for the holdoff exponent and holdoff base, in order to minimise the time it takes to complete a TW handshake.

### 3.4.1 Holdoff Exponent

In [62], it is shown that the use of a small holdoff exponent is desirable for minimising the holdoff time and $E[S]$ of a node. However, if all the nodes in a network are to make use of a small holdoff exponent, this would likely increase contention for transmission opportunities among the nodes. A consequence of this undesirable situation is that individual nodes may experience excessive delay in transmitting their data packets, as well as not being able to fairly access to

transmission opportunities. To overcome such shortcomings, it may be necessary for individual nodes to make use of different holdoff exponent values based on considerations, such as traffic service priority, operation status (which can either be base station, active subscriber station, or idle subscriber station), buffer queue size, IE transmission priority, and location in the network.

For nodes that service delay sensitive traffics, it is appropriate for them to adopt a small holdoff exponent [61, 63, 64]. This would then allow them to have quicker access to transmission opportunities. However, these nodes may become too dominant and prevent other nodes, which make use of larger holdoff exponents, from being able to access the control subframe. One way to overcome such a shortcoming is to introduce a cap on the maximum holdoff time for those nodes which service non delay sensitive traffics [63]. Once a node exceeds its specified holdoff threshold, also referred to as the virtual holdoff time, it will be allowed to compete for transmission opportunities. It is shown that by adaptively assigning holdoff exponent values to the nodes, according to the types of traffic they are currently servicing, will allow the transmission delay of Voice over Internet Protocol (VoIP) traffic to be reduced by 20%, when compared with the case that adopts uniform holdoff exponent value [63].

Also, it is possible to set the holdoff exponent value of a node according to its operation status, i.e., whether it is operating as a base station, or an active subscriber, or an idle subscriber station [64, 65]. Since a base station often has to handle a large amount of traffic, it is more appropriate for it to make use of a small holdoff exponent, in order to be able to secure more transmission opportunities for transmitting its IEs. On the other hand, active subscriber stations can set their holdoff exponent values according to the types of traffic, whether real time or non-real time that they are currently handling. Idle subscriber stations will have to make use of a large holdoff exponent, as they do not have data to transmit. Moreover, individual nodes would have to readjust their holdoff exponent values when they change their operation status [65].

However, the adjustment of the holdoff exponent based on the operation status of a given node has several shortcomings [66]. Firstly, such an approach requires a

collaborative routing protocol for determining the status of a particular node. This means that each node has to constantly refer to the routing protocol to determine whether it has been selected as a potential forwarding node for a routing path. If it is selected, the node will then change its operation status to being active and adopt a small holdoff exponent. Such a scheme is likely to incur high design and implementation complexity. Secondly, as an active node, it can request a number of frames during a single TW handshake. This suggests that an active node may not necessarily be assigned a small holdoff exponent value all the time. Thirdly, in a heavily loaded network, such as a wireless backhaul network, it is likely that all its nodes are active most of the time. As such, all the nodes will be assigned a small holdoff exponent. Consequently, the network may experience an excessive number of contentions for transmission opportunities from all the nodes, thus leading to an unnecessarily high level of collisions.

In an attempt to overcome the above shortcomings associated with selecting a holdoff exponent value based on the operation status of a given node, it is proposed in [67-69] that the buffer queue length may instead be used as a criterion for choosing the holdoff exponent value. In this case, the holdoff exponent of a given node is set to zero if its packet queue is greater than half of its buffer size. Otherwise, a unity holdoff exponent will be adopted. In the case of those less busy nodes, their holdoff exponent will be set at two. Now, if a majority of the nodes happens to have packet queues of less than half of their buffer size, the holdoff exponent values 1, 2, and 3 will be evenly distributed among all the nodes, so as to prevent them from adopting the same holdoff exponent value [69]. The use of this scheme leads to an 8 % increase in throughput and a 2 % decrease in average end-to-end transmission delay in a heavily loaded network [69]. This rather small improvement in throughput and average delay may not be considered sufficient for the adoption of such a holdoff exponent adjustment scheme in a wireless backhaul network which is expected to handle heavy traffic.

Another approach for selecting holdoff exponent value is based on the type of IEs that a given node is required to send [70]. With this approach, a node will set its holdoff exponent to zero when it is sending a confirm or grant IE. On the other hand, its holdoff exponent will be set to one when it wants to send a request IE. When there

is no IE to be sent, the node will adopt a holdoff exponent value of three. A holdoff exponent greater than three is to be avoided, as this will lead to excessive delay in transmission. This approach has also been adopted, albeit with some minor differences, in other studies as described in [71, 72]. In [71], the holdoff exponent value is set to zero if a node is intending to transmit a request or confirm IE. On the other hand, in [72], a node will adopt a holdoff exponent value of zero when it wishes to send a request or grant IE. Also, in this study, a node will continue to maintain the holdoff exponent value for the time it takes to complete its TW handshake, provided its buffer is not empty. In this way, the node will be able to make use of the same holdoff exponent to execute its next TW handshake. However, for a node which has reserved $N_F$ frames, where $1 \leq N_F \leq 128$, for data transmissions, there is no need for it to execute another TW handshake before the end of $N_F$ frames. In this case, it is more desirable for the node to raise its holdoff exponent value to give other nodes, especially those which still need to carry out TW handshake, a better chance of accessing the control subframe. As pointed out in [70], the use of an adaptive holdoff exponent based on the type of IE is not effective in enhancing throughput and delay under light traffic load conditions.

In addition, the value of holdoff exponent assigned to a node can also be determined according to its location in the network [73]. This approach is normally used in a wireless mesh network, in which all the subscriber stations are transmitting either directly or via other nodes to a base station. As such, those nodes located closer to a base station often have to relay a large amount of traffic. Under this condition, they need to receive a higher priority for accessing the control subframe to avoid traffic congestion.

The schemes described above for adjusting holdoff exponent value in a given node are more suited for networks with high node density [74]. On the other hand, for sparse networks with only a smaller number of nodes, it is more appropriate to set the holdoff exponent to zero. This, in conjunction with the default holdoff base value of four, is sufficient to allow all the nodes to have fair access to the control subframe [62].

### 3.4.2 Holdoff Base

As stated earlier in Section 3.3, the default value for the holdoff base, as specified in the IEEE 802.16 standard, is four. With this default value, a node with a holdoff exponent of zero, will holdoff for sixteen transmission opportunities after its current control message transmission. However, such a long holdoff time may not be necessary in the case of a sparse network, where the number of contentions for transmission opportunity is generally low [74]. Indeed it is shown in [74] that the use of a smaller holdoff base value, i.e., 2, could almost halve the time interval between two consecutive MSH-DSCH transmissions. Moreover, simultaneous setting of both the holdoff exponent and holdoff base to zero is to be avoided, as this could result in nodes competing with one another for every transmission opportunity. This will then lead to high control subframe wastage due to the increased likelihood of collisions [74].

It is proposed in [66] to set the holdoff exponent value for a node with a holdoff base value of zero, according to the number of its two-hop neighbouring nodes, such that

$$\text{Holdoff exponent of node } j = \lfloor \log_2 |nbr(j)| \rfloor \tag{3.6}$$

Where $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer, $nbr(j)$ is the number of nodes within the two-hop neighbourhood of a given node, $j$, including the node itself. Such an approach is considered as a static scheme. In this case, a node, which has a large number of two-hop neighbouring nodes, will make use of a large holdoff exponent to avoid a high contention rate for transmission opportunity. Conversely, a small holdoff exponent value will be adopted if it has a small number of two-hop neighbours. It is shown that the adoption of such an approach will result in an improvement in utilisation of the control subframe by a factor of 1.35, as well as a 28 % reduction in the time taken to complete a TW handshake [66].

Next, a dynamic approach is introduced in [66] for determining the holdoff exponent value for a node which is in the process of setting a data schedule. Such a node will be given a higher priority to transmit its confirm IE as soon as it has received a grant IE from the receiving node. First, the node will estimate the earliest possible

transmission opportunity for it to transmit its confirm IE, TGT_XMT_OPP, which is the next transmission opportunity after receiving the grant IE from its receiving node. This is carried out based on the scheduling information obtained via the exchanges of MSH-DSCH messages among its two-hop neighbouring nodes. A suitable holdoff exponent will then be determined according to

$$\text{Target Holdoff Exponent} = \lceil \log_2(k) \rceil \tag{3.7}$$

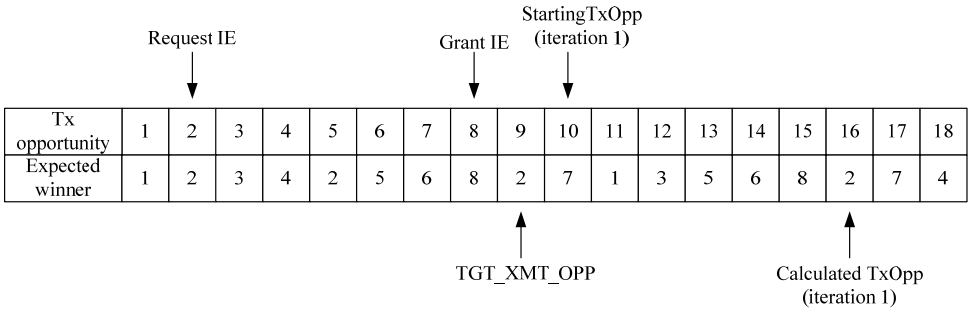
where $\lceil \bullet \rceil$ stands for rounding up to the nearest integer, $k$ is the number of transmission opportunities between the current transmission opportunity, CurrentTxOpp, of the requesting node and the next transmission opportunity, TGT_XMT_OPP. Subsequently, the first transmission opportunity in which the node will start competing for a transmission opportunity, StartingTxOpp, is determined according to

$$\text{StartingTxOpp} = \text{CurrentTxOpp} + 2^{\text{Target Holdoff Exponent}} \tag{3.8}$$

At the same time, the mesh election algorithm, as described in Section 3.4.1, will also be called upon in the determination of the next transmission opportunity for the requesting node. If the resultant next transmission opportunity, CalculatedTxOpp, is far away from the current value TGT_XMT_OPP, then an iterative approach is used to derive a more appropriate TGT_XMT_OPP value. This is done in an attempt to find the smallest exponent value that will allow a given node to transmit its confirm IE as soon as possible after it has received the grant IE. This is achieved by decrementing the Target Holdoff Exponent gradually until it yields a TGT_XMT_OPP value, which is smaller than the next transmission opportunity of the granting node. Figure 3-5 illustrates an example of steps taken in this dynamic approach.

Assume that node 2 wants to establish a data schedule with node 8. Referring to Figure 3-5, node 2 will first set its initial TGT_XMT_OPP to a value, say 9, which happens to be right after the transmission opportunity when node 8 will be likely to transmit its grant IE. From equations (3.7) and (3.8), the Target Holdoff Exponent is

calculated to be equal to 3, and the StartingTxOpp value is 10. Node 2 then uses the mesh election algorithm to determine whether it can win transmission opportunity 10. Because of the transmission opportunity contention between node 2 and its two-hop neighbouring nodes, node 2 can only manage to win transmission opportunity 16, which occurs much later than the TGT_XMT_OPP value of 9 that it initially set for itself. Hence, the Target Holdoff Exponent calculated using equation (3.7) is then transferred and becomes the Optimised Holdoff Exponent variable before the value is decremented by one. During the second iteration, node 2 is able to win transmission opportunity 9, which is still larger than or equal to the initial TGT_XMT_OPP value of 9. Therefore, the target holdoff exponent is again recorded as the Optimised Holdoff Exponent variable, and this value is decremented by one for use in the subsequent iteration. After a third iteration, node 2 manages to win transmission opportunity 5, which is less than the initial TGT_XMT_OPP of 9. At this point, the iteration process ends, and the Optimised Holdoff Exponent obtained in iteration 2 will now be used by node 2 to transmit its confirm IE in transmission opportunity 9.




(a)

| Iteration | TGT_XMT_OPP | Target Holdoff Exp | StartingTxOpp | CalculatedTxOpp | Optimised Holdoff Exp |
|---|---|---|---|---|---|
| 1 | 9 | 3 | 10 | 16 | 3 |
| 2 | 9 | 2 | 6 | 9 | 2 |
| 3 | 9 | 1 | 4 | 5 | 2 |

(b)

Figure 3-5 Illustration of a dynamic approach in arriving at a holdoff exponent: (a) the transmission opportunity straight after receiving the grant IE from the receiving node, TGT_XMT_OPP, is first estimated. The sending node then uses the mesh election algorithm to determine whether it can win TGT_XMT_OPP. However, it can only win transmission opportunity 16 in the first iteration; (b) subsequent iterations yield the Optimised Holdoff Exponent of 2.

According to the IEEE 802.16 coordinated distributed scheduling, a sending node is only able to pick up the control message transmission schedules of its two-hop neighbours via MSH-DSCH message exchanges, but not from those nodes which are two hops away from the receiving node. This scenario is shown in Figure 3-6, which indicates that the sending node is aware of the MSH-DSCH transmission schedules of those nodes, coloured in black, that are within two hops away. However, the MSH-DSCH transmission schedules of the nodes, coloured in red, are not known to the sending node. As these red-coloured nodes will also compete with the receiving node for access to the control subframe, the estimation made by the sending node regarding the transmission opportunity in which the granting node will send its grant IE, may then be inaccurate. Hence, it is possible that the holdoff exponent value obtained from the dynamic approach may yield a transmission opportunity which occurs earlier than the transmission of the grant IE. When this happens, the above dynamic approach will fail to minimise the time it takes to complete a TW handshake.

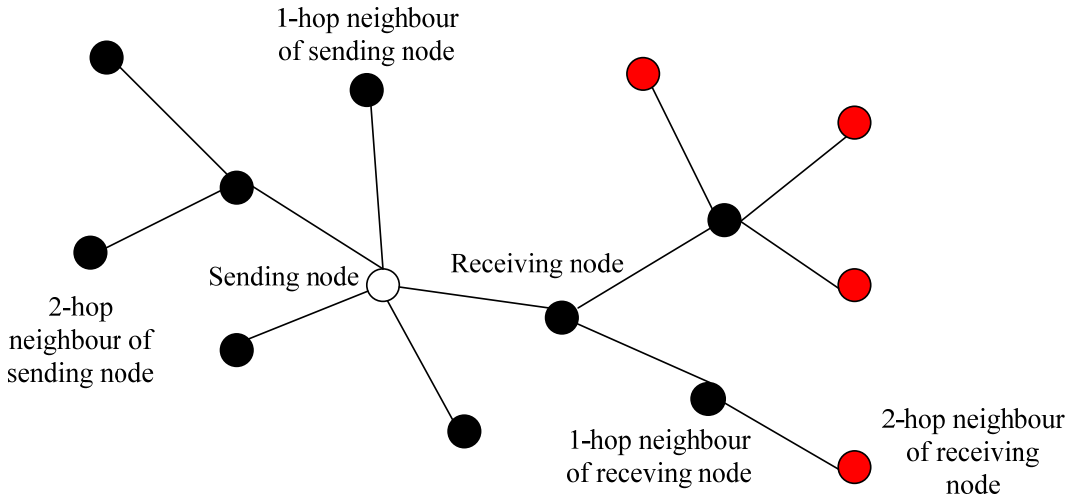



Figure 3-6 The sending node does not know the MSH-DSCH transmission schedule of nodes, coloured in red, which are two-hop away from the receiving nodes.

### 3.4.3 Data Minislot Allocation

During a TW handshake, it is possible that a sending node may fail to get any minislots granted by the receiving node. Under such a situation, the sending node is

likely to be restrained from data transmission for a long period of time. This can result in throughput degradation and large transmission delay. For this reason, it is important to examine the data minislot allocation process in a TW handshake.

One simple approach is to distribute the available minislots equally among all the network nodes [75]. With this approach, the number of minislots allocated to node *i*, $N_i^s$ becomes

$$N_i^s = \left\lfloor \frac{N_{fr}^s}{N} \right\rfloor \tag{3.9}$$

where $N_{fr}^s$ is the total number of minislots in a frame, $N$ is the number of nodes, and $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer.

Another way is to allocate minislots to a node according to its number of data flows. Here, a data flow refers to all the data transmitted between a given pair of source and destination nodes [75]. This also means that a larger number of minislots will be allocated to those network nodes which happen to carry large data flows, such as in the case of a base station in a wireless backhaul network. Consequently, the number of minislots allocated to a node is given by

$$N_i^s = \left\lfloor \frac{N_i^{fl} N_{fr}^s}{\sum_{k=1}^{M} N_k^{fl}} \right\rfloor \tag{3.10}$$

where $N_i^{fl}$ and $N_k^{fl}$ is the number of flows for the node *i* and node *k*, respectively. However, it is possible that the actual traffic load of a given node may be high, even though its number of data flows is low. This suggests that the number of data flows of a node may not necessary reflect its true traffic load condition.

In an attempt to overcome the shortcoming of the above flow proportional scheme in allocating channel resources, a traffic load dependent minislot allocation scheme is

proposed in [75, 76]. With this scheme, a larger number of minislots is allocated to a node with higher traffic demand, which is reflected through its number of minislots requested.

Besides having to determine the appropriate number of minislots for a node, it is also necessary to decide on the number of frames that can be requested by the node. This is essential for efficient utilisation of data subframe that leads to improved network throughput [77]. In [77], it is proposed that the number of minislots, $|Dx|$ to be allocated is calculated according to

$$|Dx| = \left(T_F - 7 \times SymbolTime \times MSH\_CTRL\_LEN\right) / M \tag{3.11}$$

where $T_F$ is the frame duration, *SymbolTime* is the duration of an OFDM symbol in the physical layer, MSH-CTRL-LEN is the number of control message transmission opportunities in a control subframe, and $M$ is the number of transmissions within a two-hop neighbourhood. Also, the number of frames that can be allocated to a node is given by

$$n = \frac{R_D T_{handshaking}}{|Dx| R_{PHY} - R_D T_F} \tag{3.12}$$

where $R_D$ is the data generating rate, $T_{handshaking}$ is the time needed to complete a three-way handshake, and $R_{PHY}$ is the physical layer data rate. The values of $R_{PHY}$ and $T_F$ are generally considered as constants, so that $R_D$ and $T_{handshaking}$ can be computed, such that

$$R_D = R_D^{'}\alpha + R_D\left(1-\alpha\right) \tag{3.13}$$

$$T_{handshaking} = T_{handshaking}^{'}\beta + T_{handshaking}\left(1-\beta\right) \tag{3.14}$$

where $\alpha$ and $\beta$ are tuning parameters used in the estimations of $R_D$ and $T_{handshaking}$ respectively. With considerations given to both fairness and spatial reuse, $|Dx|$ may be modified to become

$$|Dx|_{candidate} = \begin{cases} |Dx| & if\ \left(|Dx| \geq |Dx|_{\min}\right) \\ \left(|Dx|_{\min} + |Dx|\right)/2 & if\ \left(|Dx| < |Dx|_{\min}\right) \end{cases} \quad (3.15)$$

where $|Dx|_{\min}$ is the minimum number of minislots required by a given node. It is shown that after these modifications, the packet loss rate can be reduced by 26 % while the minislots utilisation is improved by 19 % under a heavy load condition, when compared to the case that adopts random scheduling [77]. However, there is no guarantee that the numbers of minislots and frames calculated based on the above approach will be available and granted by the receiving node.

In case a receiving node does not have a sufficient number of minislots and frames requested by a sending node, it may adjust and allocate resource within its available capacity [78]. However, the receiving node is only able to do such an adjustment if the reserved bit in the received MSH-DSCH request IE is set at one. Referring to the example of Figure 3-7, nodes A and B send a separate request to the same granter, node C. Assume that node A requests 128 frames with two minislots in each frame starting from Frame i. At about the same time, node B requests to transmit in two minislots in Frame i.

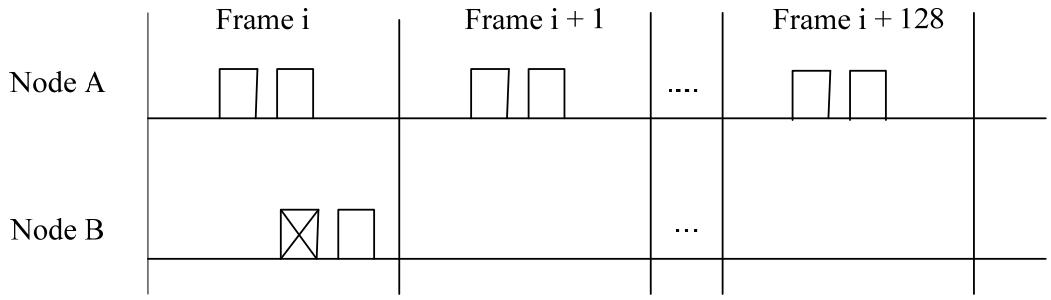



Figure 3-7 Node A and node B send a separate request to node C. However, one of the two requested minislots of node B overlaps in time with that of node A in Frame i.

From Figure 3-7, it can be observed that one of the two requested minislots of node B overlaps in time with that of node A in Frame i. Let assume that the request of node A arrived first at node C. In this case, Node C will grant the request of node A, and reject the one from node B. Now, if node C were allowed to adjust the number of

frames requested by node B, it can grant node B two frames, each with one instead of two minislots as originally requested, as shown in Figure 3-8. In this way, node B will be able to transmit data, albeit at a lower rate, immediately after it has completed its TW handshake.

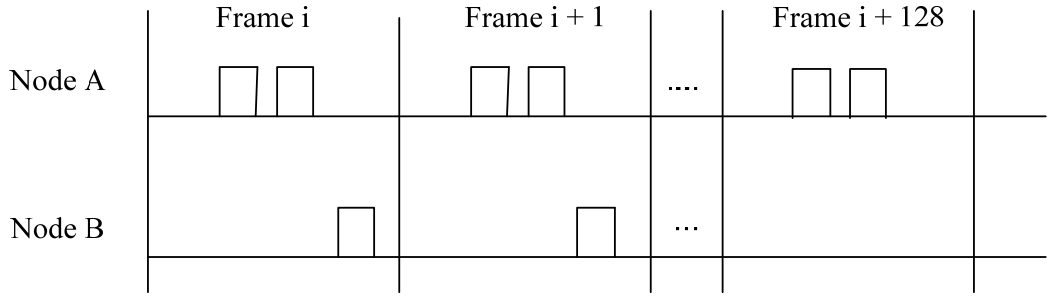



Figure 3-8 Node B is granted with two frames, each with one minislot after node C is allowed to adjust the number of frames requested by node B.

The above is referred to as the demand persistence adjustment scheme. It allows a node whose requested resource would otherwise not be granted to receive a larger number of frames, each with less minislots, to make up the total minislots requested by the node. However, it would take the node longer to complete sending all its data. According to Figure 3-7, node B initially would like to transmit all its data within a single frame. Moreover, with its requested resource being reduced by the granting node, it would now take two frames for node B to complete its data transmission, as shown in Figure 3-8. With the adoption of such an approach, it is shown that the average end-to-end delay may be reduced by only 3 % [78].

An alternative approach, called the multi-grant scheme, has been proposed in [79] for overcoming the conflict situation, as illustrated in Figure 3-7, between the requests of node A and node B. Now, instead of having to allocate resources only in the form of consecutive minislots, as specified by the original IEEE 802.16 TW handshake process, the multi-grant scheme enables a receiving node to grant multiple non-consecutive minislots in a frame until one of the following conditions is reached.

1. The requested number of minislots has been fulfilled.
2. No requested minislot is available.

3. The number of allocations has reached the pre-defined threshold, $G_{Thres}$, which is introduced to prevent a node from monopolising the network bandwidth.

Now, refer again to the previous conflict situation as depicted in Figure 3-7. With the above multi-grant scheme, node C is able to allocate two unoccupied non-consecutive minislots in Frame i to node B, as illustrated in Figure 3-9. With this allocation, node B will be able to complete its data transmission in one frame rather than two as for the case of Figure 3-8. As a result, the multi-grant scheme is shown to yield a 45 % increase in throughput, as well as a 53 % reduction in average transmission delay [79].

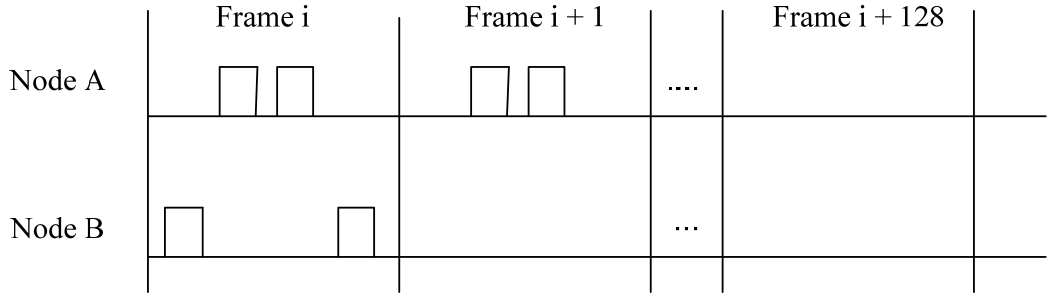



Figure 3-9 With the use of the multi-grant scheme, node C is able to allocate two unoccupied non-consecutive minislots in Frame i to node B.

Furthermore, studies described in [80-82] have identified that a request conflict may also occur due to the hidden node problem. This occurs when two receiving nodes, located beyond the coverage of one another, happen to grant the same set of minislots to their respective transmitting nodes. This scenario is illustrated in Figure 3-10. In this case, nodes P and R send their requests to nodes Q and S, respectively. Node S, being two-hops away from node Q, is unaware of the minislots granted by node Q to node P, also selects the same set of minislots to be granted to node R. When node R receives the grant from node S, it also overhears the grant from node Q to node P. Node R then decides that there is a conflict in the minislots allocation, which may lead to a collision at node Q. Consequently, node R restrains itself from confirming the grant from node S. Such a situation is known as grant-withdrawal.

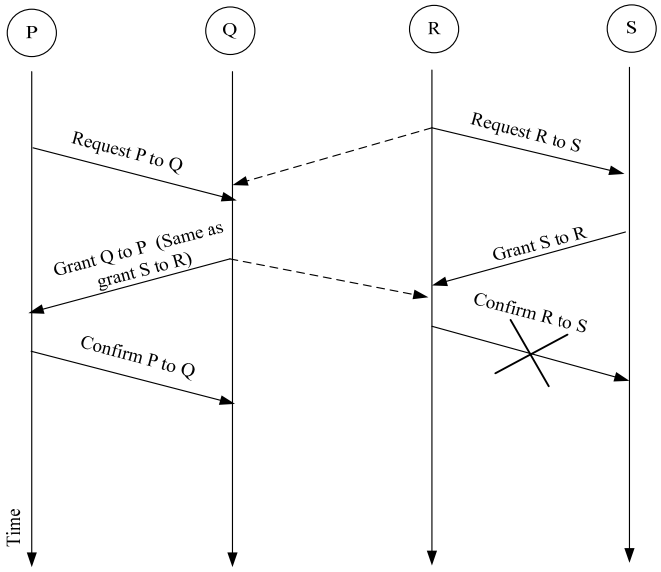



Figure 3-10 After node R receives the grant from node S and overhears the same grant from node Q to node P, it detects a minislot allocation conflict, which restrains it from confirming the grant from node S.

Several studies have attempted to overcome the above grant-withdrawal scenario caused by the hidden node problem, which in this case arises between nodes Q and S [80, 81, 83, 84]. One way is to adopt a regranting procedure, as shown in Figure 3-11, so that node S will be able to grant node R a new set of minislots, after it detects a handshake failure. In this way, node R is able to confirm the grant of minislots and continue to transmit data to node S. However, the grant-withdrawal scenario occurs only when nodes do not transmit an availability IE during their requests for bandwidth [80, 81]. As with the example illustrated in Figure 3-12, if both nodes P and R were to transmit their requests for channel resources in conjunction with the use of availability IEs, then node Q will be able to detect a request conflict. As a result, node Q will ignore the request from node P, thus avoiding the grant-withdrawal scenario. Moreover, this also suggests that the regranting scheme will not be able to solve the hidden node problem. Under this situation, the hidden node problem may be overcome through the use of multiple channels, and this will be further discussed in Section 3.4.4.

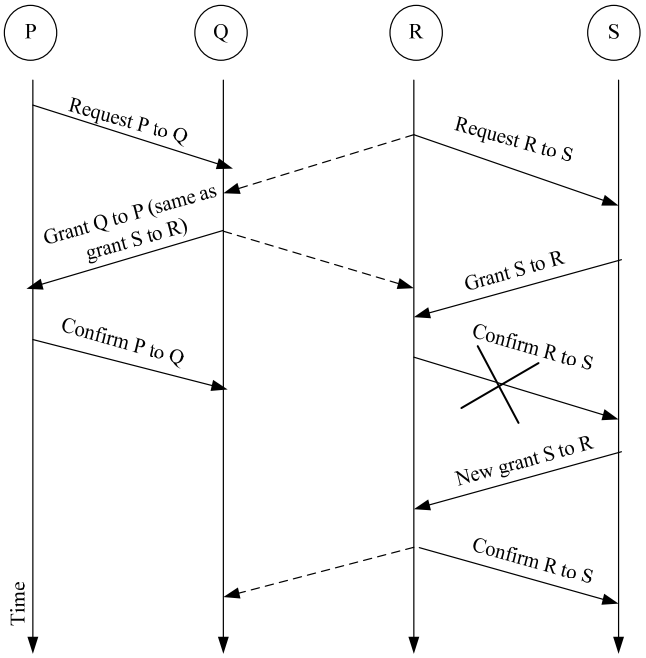



Figure 3-11 Regranting scheme allows node S to grant a new set of minislots to node R.

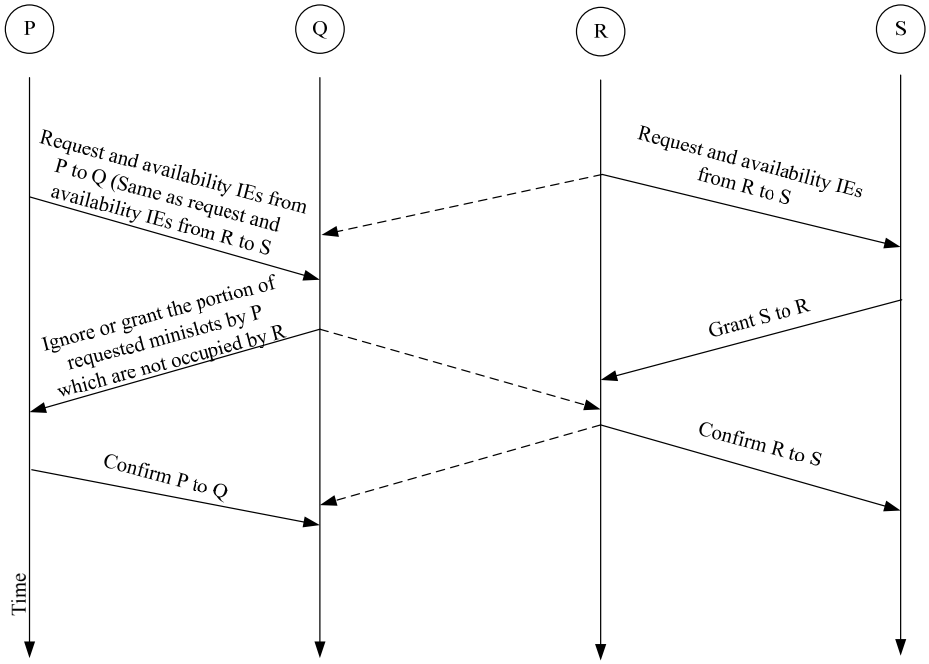



Figure 3-12 Hidden node scenario when the availability IE is used during a three-way handshake.

### 3.4.4 Use of Multiple Channels in IEEE 802.16 Coordinated Distributed Scheduling

In a wireless network, co-channel interference occurs at a given node when it receives signals simultaneously from more than one sending node. One likely cause of this situation is attributed to the hidden node problem. It is well known that the above co-channel interference can be avoided if each of the sending nodes employs a different frequency channel. Again, consider the example of Figure 3-10, which shows nodes Q and S grant the same minislots to nodes P and R, respectively. This time, collisions will not occur at node Q, if node P and node R make use of separate channels to transmit their data.

There are two ways of implementing multiple frequency channels in a wireless network, either equip a node with a single transceiver or multiple transceivers. For the former, the node can switch between channels from time-to-time if necessary, but this will incur unavoidable delay due to the settling time of the frequency synthesiser. On the other hand, when a node is equipped with multiple transceivers, it can transmit and receive simultaneously in more than one frequency channel without the need to switch between frequencies.

In an IEEE 802.16 wireless backhaul network, a node cannot transmit concurrently with its one-hop and two-hop neighbours operating in the same frequency channel. This can only be overcome, if different frequency channels are assigned to these one-hop and two-hop neighbours. The IEEE 802.16 standard specifies sixteen logical frequency channels for multi-channel transmissions [84]. Moreover, most current research effort tends to focus on assigning frequency channels to network nodes via centralised scheduling schemes [85-93]. With these schemes, a base station is responsible for allocating frequency channels centrally to all the individual subscriber stations.

When operating with distributed scheduling, channel assignment is carried out through exchanges of control messages among two-hop neighbouring nodes. For example, in a wireless network, which equips its individual nodes with a single radio transceiver, all the nodes have to tune to a common frequency channel, for instance,

Channel 1, as illustrated in Figure 3-13, for the exchange of control messages. After that, each node will then be able to choose a different frequency channel for data transmission under the direction of a distributed channel assignment scheme [80, 81, 84]. With this approach, each node has to record the status of each minislot, frame and channel in a three-dimensional bit map, as shown in Figure 3-14.

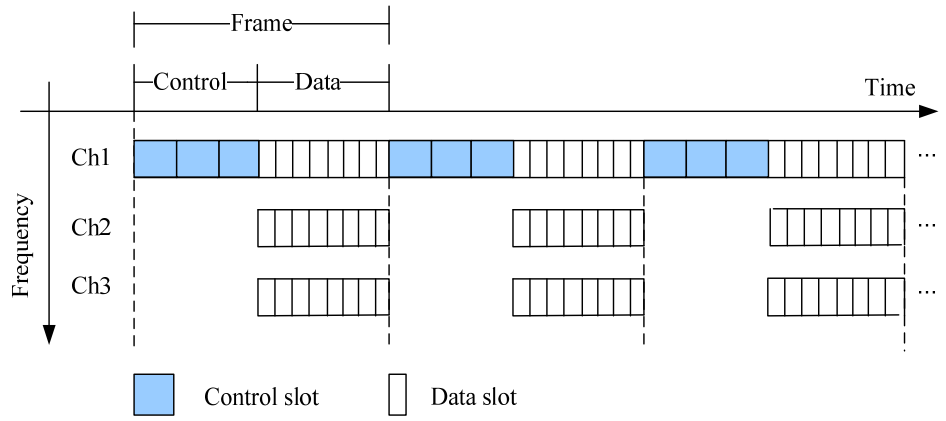



Figure 3-13 An example of frame structure involving three frequency channels.

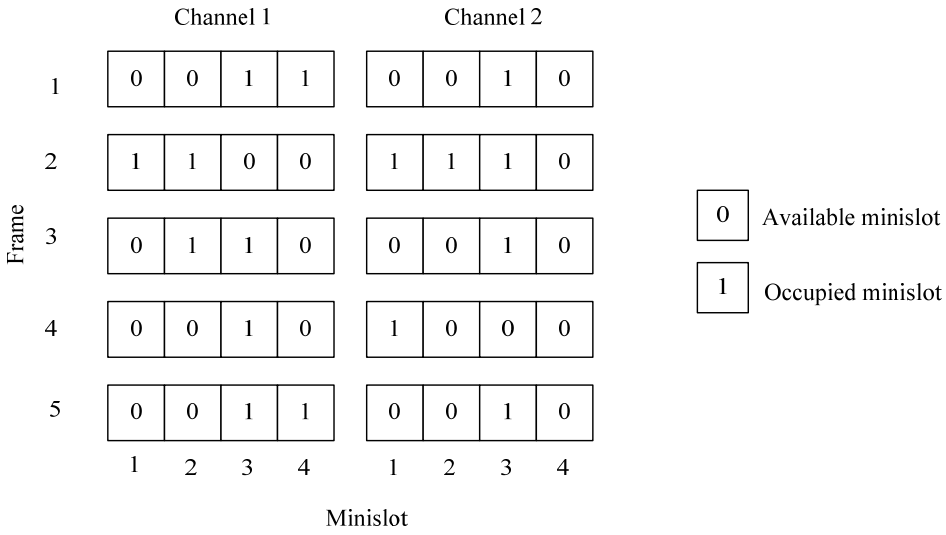



Figure 3-14 A three-dimentional bit map is used to record the status of each minislot, frame and channel. In this bit map, an available minislot is indicated using a logical "0" bit and logical "1" is used to denote an occupied minislot.

The bit map associated with a given node records not only its own grants and confirmations but also the grants and confirmations of its neighbours, even though the node itself is not involved in the negotiation for bandwidth allocation. This bit

map is used by a receiving node to determine whether the particular minislots requested by a sending node are available. A minislot in a particular frame and channel is said to be not available for bandwidth grant in the presence of any one of the following conditions:

1. The receiving node transmits or receives in the same minislot and frame.
2. The sending node transmits or receives in the same minislot and frame.
3. One of the neighbours of the sending node happens to transmit using the same minislot, frame, and channel.

To search for free minislots in the bit map, a node will first randomly select a frequency channel. This step is necessary in order to reduce the likelihood that a given node and its two-hop neighbour, that are out of the interference range of one another, grant the same minislots to the node located in between them. The node then decides upon the range of frames for which it is allowed to grant minislots. This range of frames is referred as the schedule horizon. Once the schedule horizon is established, the node will begin its search for available minislots starting from the first minislot in the first frame of the schedule horizon. If the first frame of the schedule horizon in the selected channel does not have free minislots, the search will continue in a different channel. In the event of no free minislots being available from the same frame in all the channels, the node will then start to search in the second frame. This process continues until the bandwidth requested by the sending node is fulfilled or the search has checked all the minislots in all the frames and channels. This process is illustrated in Figure 3-15.

Let assume that a given receiving node has to allocate three minislots to a sending node. In doing so, it begins the search routine by first randomly selecting one of the two possible frequency channels, say channel 1 in the example shown in Figure 3-15. It then starts to search for free minislots in the first frame of the schedule horizon, i.e., frame 2. Figure 3-15 shows that minislots 3 and 4 of frame 2 in channel 1 are available. Hence, these two minislots are selected and the node will in turn send a grant IE back to the sending node. Since not all the request of the sending node has

been met, the node will continue to search for further free minislots in channel 2. Although minislot 4 in channel 2 is available, it overlaps in time with the minislot 4 in channel 1. As such, it is not a valid resource to be considered for allocation. The reason is that a given node cannot operate on two separate channels simultaneously using a single transceiver. Once the availability of all the minislots in both channels in frame 2 are checked, the node will then continue its search in the second frame of the schedule horizon, i.e., frame 3 in this example. As minislot 1 of frame 3 in channel 1 is available, and does not overlap in time with minislots 3 and 4 in frame 2 in channel 1, the node will include this minislot for allocation by sending another grant IE back to the sending node. At this point, the entire amount of resource requested by the sending node has been met, so this ends the search process.

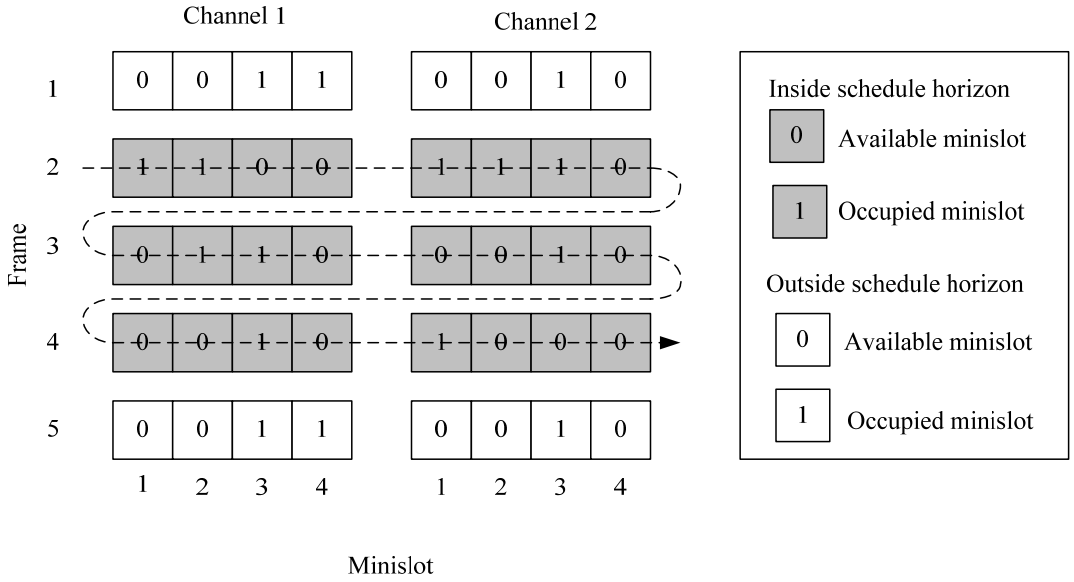



Figure 3-15 Searching for free minislots in a three-dimensional bit map associated with a network node equipped with a single radio transceiver operating with multiple frequencies. The highlighted boxes represent the schedule horizon where the search is to be carried out.

## 3.5 SUMMARY

With its ability to support single-hop over a range of up to 50 km as well as multi-hop wideband radio transmissions, the IEEE 802.16 wireless communication technology is attractive for application in wireless backhaul networks. The IEEE

802.16 standard defines the specifications for both the physical and MAC layers. Also, the standard specifies two operating modes, namely, PMP and mesh. Yet, only mesh mode supports multi-hop transmissions, which are crucial for wireless backhaul networks. When operating in the mesh mode, transmission in the physical layer is carried out using OFDM while the MAC layer is based on TDMA. In order to coordinate the transmissions of individual nodes in mesh mode, three scheduling schemes are defined, i.e., centralised scheduling, coordinated distributed scheduling, and uncoordinated distributed scheduling. For a wireless backhaul network, coordinated distributed scheduling offers a better option by requiring substantially shorter connection setup time, as well as ensuring collision-free control message transmissions [61]. This scheduling scheme makes use of a three-way (TW) handshake procedure to establish data transmission schedules between two nodes. As a node can only transmit data after it has completed a TW handshake, the time taken for this handshaking procedure plays a vital role in determining the achievable overall network throughput and transmission delay.

The time for completing a TW handshake is influenced by the holdoff time, and time lost during the contention for a transmission opportunity in the control subframe, $E[S]$ [61]. Both the holdoff time and $E[S]$ are governed by two parameters, namely, the holdoff exponent and holdoff base values. The former is adjusted based on specific criteria, which in turn depend on factors such as traffic service priority, operation status, buffer queue size, IE transmission priority, and the location of a node. For a network involving a large number of nodes constantly contending for transmission opportunities, it is more appropriate for these nodes to be assigned with different holdoff exponents according to their individual circumstances.

For a sparse network, such as a wireless backhaul, the rate of contentions for transmission opportunities is relatively low. Hence, it is possible to make use of a small holdoff exponent in order to minimise the time needed to complete a TW handshake. Moreover, according to [74], the use of the default holdoff base value of 4, specified in the IEEE 802.16 standard, is not optimal and can give rise to excessive transmission delay. It is proposed in [66] that this may be corrected by adopting a holdoff base value according to the network node density. It is shown that

the adoption of such an approach can double the control subframe utilisation and reduce the duration of a TW handshake by 28 % [66].

During a TW handshake, it is possible that a node may fail to obtain a minislot for data transmission. As a result, the node is restrained from transmission. In order to minimise the rate of occurrences of such a situation, it becomes essential to search for better ways of allocating data minislots. The actual ways that could be adopted for data minislots allocation are not specified in the IEEE 802.16 standard. Several research studies have been devoted to look at possible ways to carry out data minislots allocation, and these include fair allocation among all the network nodes [75], flow proportional minislots allocation [75], and traffic load dependent minislots allocation [75, 76]. Moreover, possible conflicts may occur during minislot requests between neighbouring nodes due to the so called hidden node problem. Several schemes have been proposed to overcome such conflicts, including the adjustment to the number of requested frames [78], multi-grant of minislots [79], and regranting minislots [80, 81]. However, the first of these schemes tends to increase the time needed for a node to complete its data transmission. On the other hand, the multi-grant scheme involves an increase in the overhead of control messages [94]. As for the regranting scheme, it fails to overcome conflicts involving minislot requests between neighbouring nodes when availability IE is used during a TW handshake. This calls for a more effective way to resolve minislot request conflicts and this will be further discussed in Chapters 4 and 5. Also, the use of multiple frequency channels in overcoming such a conflict will be investigated in Chapter 6.

# CHAPTER 4

# DESIGN OF A FAILURE RESILIENT WIRELESS BACKHAUL

## 4.1 INTRODUCTION

All the large population centres of the world are today interconnected through a complex web of telecommunication backbone infrastructure. However, many regions of low population density, particularly those in the outlying areas far away from metropolitan centres, remain without access to modern information and communication technology (ICT) services. It is envisaged that a multi-hop IEEE 802.16 wireless backhaul could offer a flexible and cost-effective solution for these distant remote community centres to access modern broadband services gateways. In order to avoid a possible disruption of telecommunication services in the presence of an occasional link or node failure, it is necessary to incorporate alternative paths into the network for rerouting traffic. However, the use of extra nodes and links to establish the alternative paths requires careful planning so that the specified failure resilience is able to be achieved cost effectively.

In this chapter, a simple failure resilient multi-hop IEEE 802.16 wireless backhaul network for connecting a remote community to a gateway node located in the regional or metropolitan centre is proposed. The various factors considered in the design of such a network are first described in Section 4.2. Next, the proposed failure resilient IEEE 802.16 wireless backhaul topology is presented in Section 4.3. Then, the performance of the proposed topology operating according to the IEEE 802.16 coordinated distributed scheduling standard is evaluated in Section 4.4.1, and the simulation results obtained are presented in Section 4.4.2. The problem encountered with the coordinated distributed scheduling is then examined in Section 4.4.3 and this is followed by the description of a new scheme proposed for overcoming such a

problem in Section 4.4.4. The performance of the new scheme is then analysed in Section 4.4.5. Lastly, Section 4.5 summarises the chapter.

## 4.2 DESIGN FACTORS

Most published works referred to in Chapter 2 are devoted to the design of failure resilient networks in urban or suburban areas. In such networks, the locations of the nodes are first determined based on certain specified criteria, such as the subscriber population distribution and user traffic demand. Communication links are then introduced to connect the various nodes in a manner which will meet the specified network requirements, such as failure survivability level and minimum establishment cost.

On the other hand, the design of a failure resilient multi-hop IEEE 802.16 wireless backhaul network for delivering broadband services from a metropolitan centre to a remote community tends to involve the use of long routes. The nodes along these long routes often have little or no users around and they merely serve to relay traffic. Under this condition, the locations of the nodes are more likely to be determined based on the geographical terrains encountered. In an attempt to simplify the design problem, it is assumed that the geographical terrain between the remote community and the metropolitan centre is homogenous. This then leaves us to consider other factors which are likely to influence the design, such as the network cost, failure scenario, level of connectivity, interference, traffic rerouting strategy, and transmission delay during rerouting, to determine the suitable location of the nodes, the number of nodes to cover the distance between the two community centres, and also the interconnection links among the network nodes. These factors are discussed in Sections 4.2.1 to 4.2.4.

### 4.2.1 Network Cost

As mentioned in Section 2.4.1, the deployment cost of a backhaul network is mainly attributed to capital equipment and labour costs. These costs are highly variable, depending on equipment vendors and the inherent geographical terrains involved [7,

47]. Nonetheless, it is reasonable to assume that the deployment cost could be minimised by employing as few base stations as possible to satisfy the required network performance. As such, a chain topology, as shown in Figure 4-1, is a viable option for connecting two distant communities, i.e., Communities X and Y, with the least number of base stations.

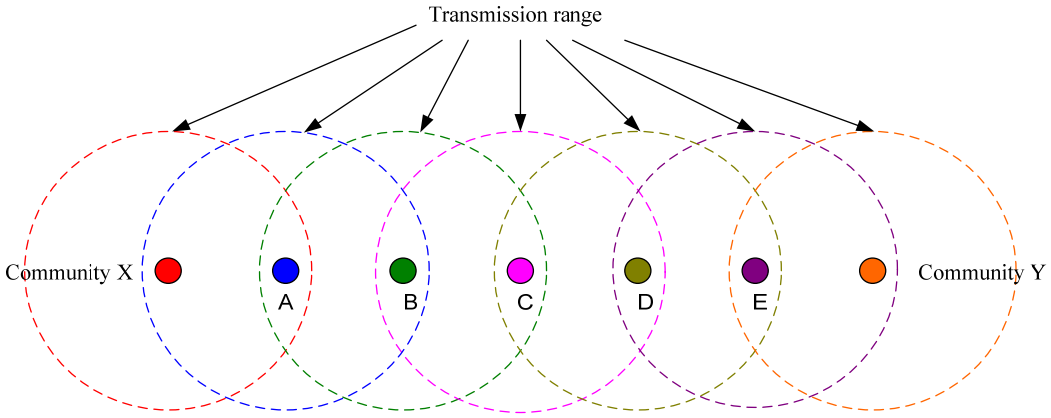



Figure 4-1 A chain topology used for connecting two distant communities, X and Y. Each dotted circle represents the coverage of the base station located in its center.

### 4.2.2 Failure Scenarios

When the simple chain topology of Figure 4-1 is adopted for the backhaul network, communications between Communities X and Y will be disrupted in the event of any single node or link failure, as shown in Figure 4-2.

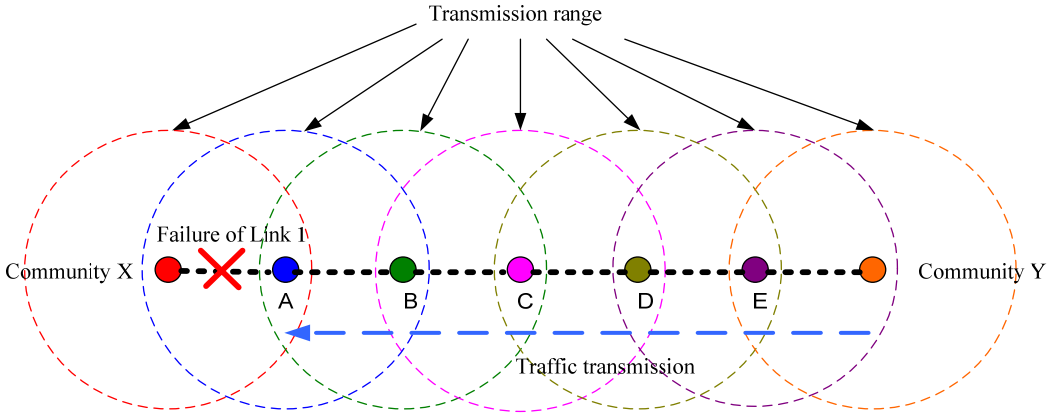



Figure 4-2 Traffic from Community Y fails to arrive at Community X due to the failure of Link 1.

An obvious way of overcoming this shortcoming is to duplicate the chain of nodes to provide an additional separate transmission path between the two communities, as shown in Figure 4-3(a). Although the resultant two parallel path backhaul network is more failure resilient, communications between Communities X and Y will still fail if both branches suffer from a single node or link failure as depicted in Figure 4-3(b). If required, additional chains of nodes may be introduced to enhance failure resilience of the backhaul network, but this will inevitably lead to excessive network deployment cost.

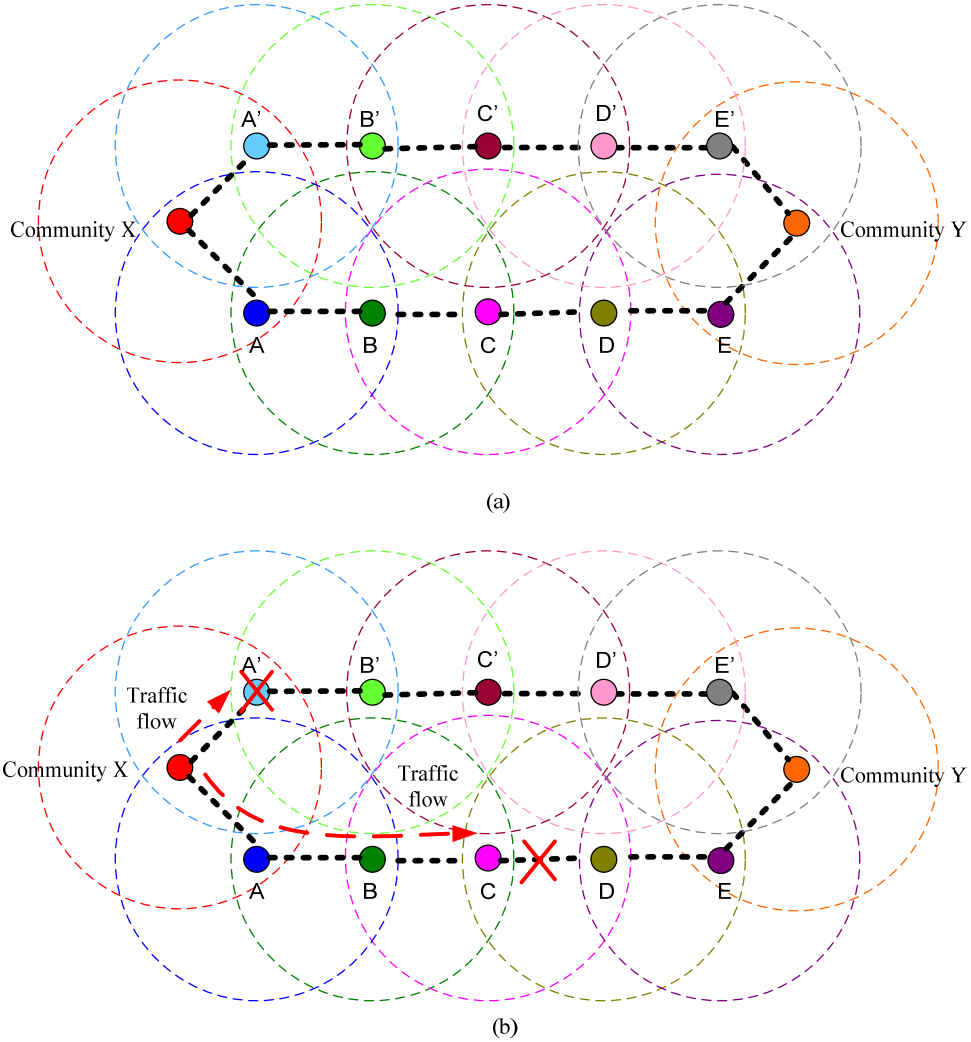



Figure 4-3 Two parallel links are used to connect two communities, X and Y: (a) the network topology; (b) failure of any link or node at both paths will disrupt the data transmission between the communities.

Instead, a single chain network may be made to better withstand link and node failures by having specific alternative paths introduced to bypass the faulty link and node. For example, in Figure 4-4(a), an extra node, A', is added to the network. It connects with both Community X and node A to provide two possible paths between Community X and node A. The new path and the original Link 1 are link-disjointed from one another, i.e., they do not share a common link.

However, the above solution is not effective in overcoming transmission disruption due to a node failure. In this case, extra nodes and links are needed to establish a node disjoint path for rerouting traffic away from the faulty node. According to the example of Figure 4-4(b), a second node, B', is added to combine with node A' in forming a node disjointed path, which bypasses the failed node A.

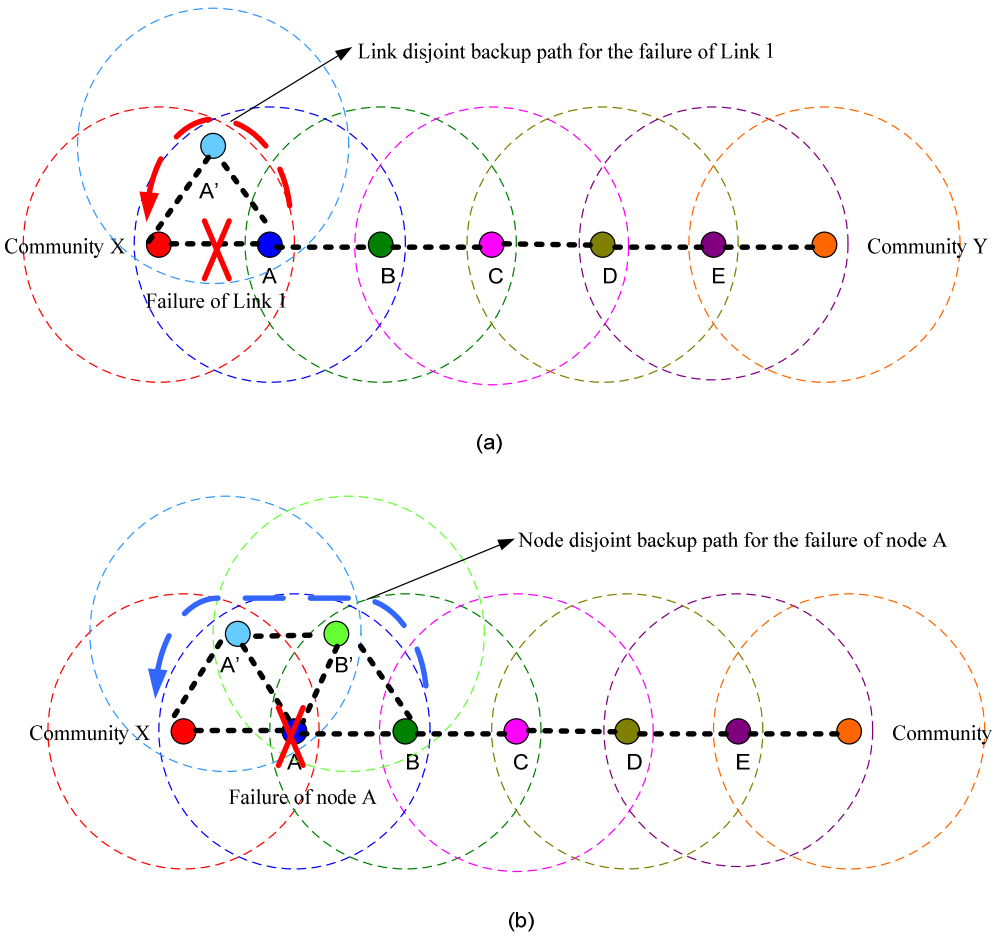



Figure 4-4 Backup paths: (a) link disjoint paths between community X and node A; (b) node disjoint paths to provide backup for the failure of node A.

In an attempt to keep the deployment cost low, only minimum number of extra links and nodes are introduced. For this reason, it is essential to be able to make use of these extra links and nodes in such a way that they could be shared in protecting different faulty links or nodes. As depicted in Figure 4-5, nodes A', B', and C' are used to provide a shared backup path for rerouting traffic during a failure which could occur at either node A or node B.

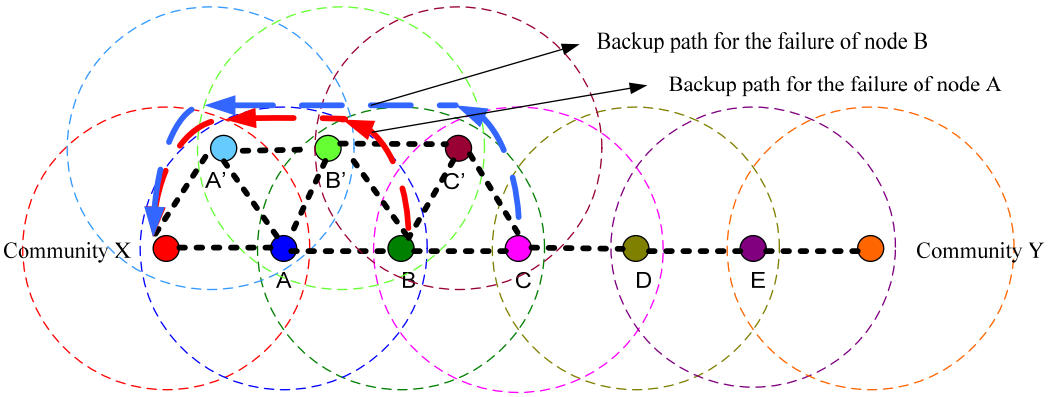



Figure 4-5 Node A', B', and C' are used to establish a shared backup path for rerouting traffic when node A or B fails.

### 4.2.3 Level of Connectivity and Interference

As a wireless backhaul network is used to serve a large population of users, it should continue to operate even in the event of multiple link or node failures. Such a stringent requirement usually means that each link and node would have to be protected by providing at least two or more backup paths. In other words, a given node will have to be able to communicate with three or more neighbouring nodes. Now, if these nodes are equipped with omnidirectional antennas and operating on the same frequency, they will give rise to mutual co-channel interference. In order to keep the interference level to a manageable level, it is desirable for each node to have not more than three neighbours. Again, refer to the network of Figure 4-5, which shows that node A is connected to four neighbouring nodes, i.e., nodes X, A', B, and B'. Moreover, we could rearrange the positions of individual nodes in the network, as illustrated in Figure 4-6, to limit the number of neighbours for a given node to not more than three.

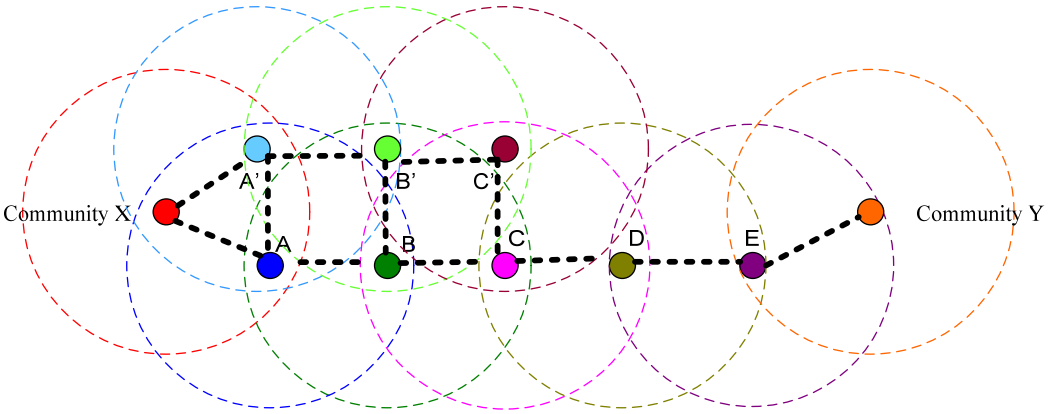



Figure 4-6 The positions of the individual nodes are rearranged to limit the number of neighbouring nodes for a given node to not more than three in order to reduce the co-channel interference.

### 4.2.4 Traffic Rerouting Strategy and Transmission Delay

As discussed in Section 2.4.4, the local rerouting strategy is rather simple to implement and the rerouted traffic is likely to reach its destination with less delay. Such a rerouting strategy is, therefore, attractive for use in a wireless backhaul network incorporated with alternative transmission paths. An example of local rerouting is shown in Figure 4-7 with the link between nodes B and C failed.

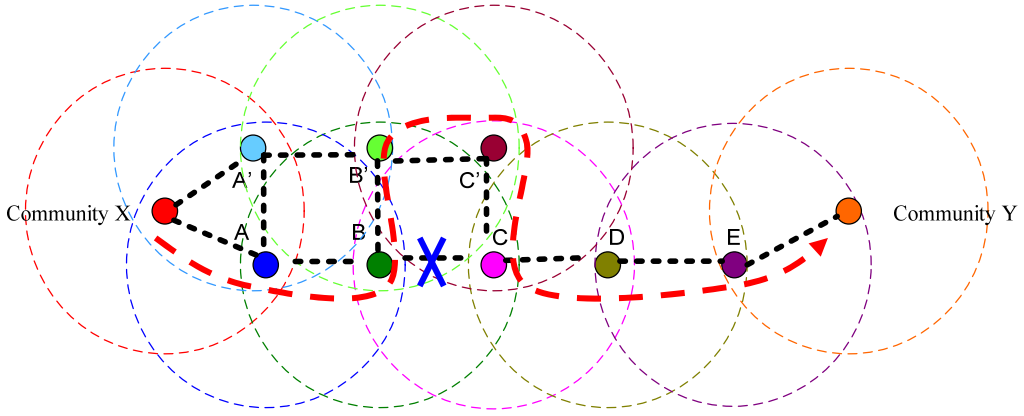



Figure 4-7 Local rerouting makes use of the alternative path established from node B when the link between node B and C fails.

In this case, an alternative path is established from the node immediately before the failed link, i.e., node B, to reroute traffic away from the failed link to travel on the alternative path provided by nodes B, B', C', and C. Also, this alternative path involves only two additional hops compared with the original direct route.

Consequently, the additional transmission delay incurred for the rerouted traffic is kept to an acceptable level.

## 4.3 FAILURE RESILIENT TOPOLOGY

After considering the various design factors discussed in Section 4.2, a relatively simple ladder topology is proposed for the implementation of a failure resilient wireless backhaul. An example of a six-hop wireless backhaul network connecting two distant communities, X and Y, is shown in Figure 4-8.

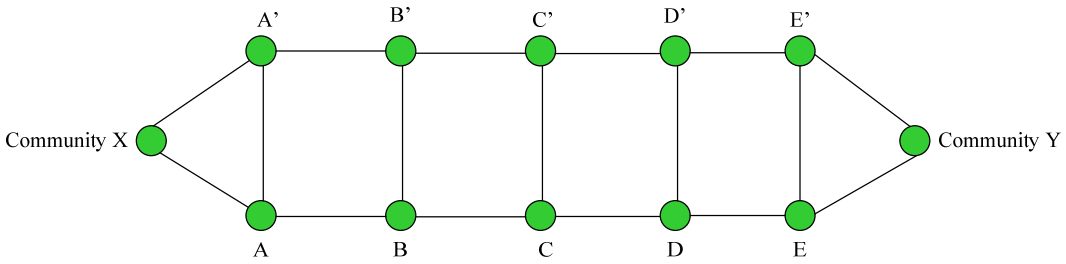



Figure 4-8 A six-hop ladder topology connecting Communities X and Y.

This failure resilient ladder topology consists of two chains of relay nodes, namely A, B, C, D, E, and A', B', C', D', E', which serve the gateway nodes, X and Y, at two distant communities. Each chain is established using the minimum number of nodes and links to cover the distance between nodes X and Y. Although only one single chain of relay nodes is sufficient to form the wireless backhaul, its operation will be disrupted in the presence of a single node or link failure. Through the use of an additional chain of relay nodes, it is possible to provide each individual node and link of the backhaul with at least one backup path. Such an arrangement is likely to incur minimum network deployment cost as it only requires the minimum number of additional nodes to realise the necessary backup paths. Furthermore, the two chains of relay nodes provide the two shortest direct transmission paths for relaying traffic between the two communities, X and Y, during normal operation.

On the other hand, the cross links between each pair of nodes, located at the same position along each of the two chains, for example, the link between node A and A', is used to locally reroute traffic bypassing a faulty link or node. The proposed ladder

topology is able to sustain multiple link and node failures. In the case of *n* failures in the network, the rerouted traffic only has to traverse at most *n* additional hops compared with travelling by the direct path. Figure 4-9 shows that traffic from Community X can be rerouted to arrive at its destined Community Y even though both nodes C and E' were to fail simultaneously.

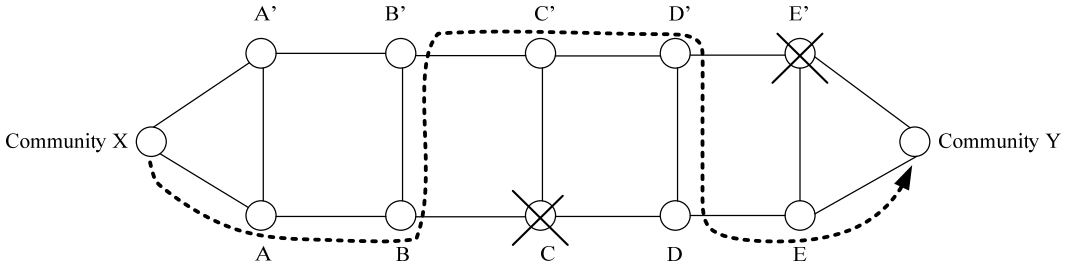



Figure 4-9 The wireless ladder backhaul can survive two simultaneous failures.

However, the proposed ladder topology will not be able to overcome the following three scenarios involving multiple failures:

1. Concurrent failures of the two nodes responsible for a cross link, such as nodes B and B' as shown in Figure 4-10(a).
2. Simultaneous failures of the links occurring at the same hop level in each of the two network branches. An example of this scenario is the link failures between nodes B' and C', and nodes B and C, as shown in Figure 4-10(b).
3. Failures of two nodes, which occur in consecutive hop sequence across the two branches of the ladder network. This scenario is typified by the failures of nodes B' and C, as shown in Figure 4-10(c).

When compared with single failure events, the likelihood of simultaneous occurrences of multiple failures is substantially lower [25, 26]. These observations, when taken in conjunction with the desire to keep the deployment cost of the network to an acceptable level, suggest that the need for the network to sustain all possible multiple failure scenarios may be somewhat relaxed.

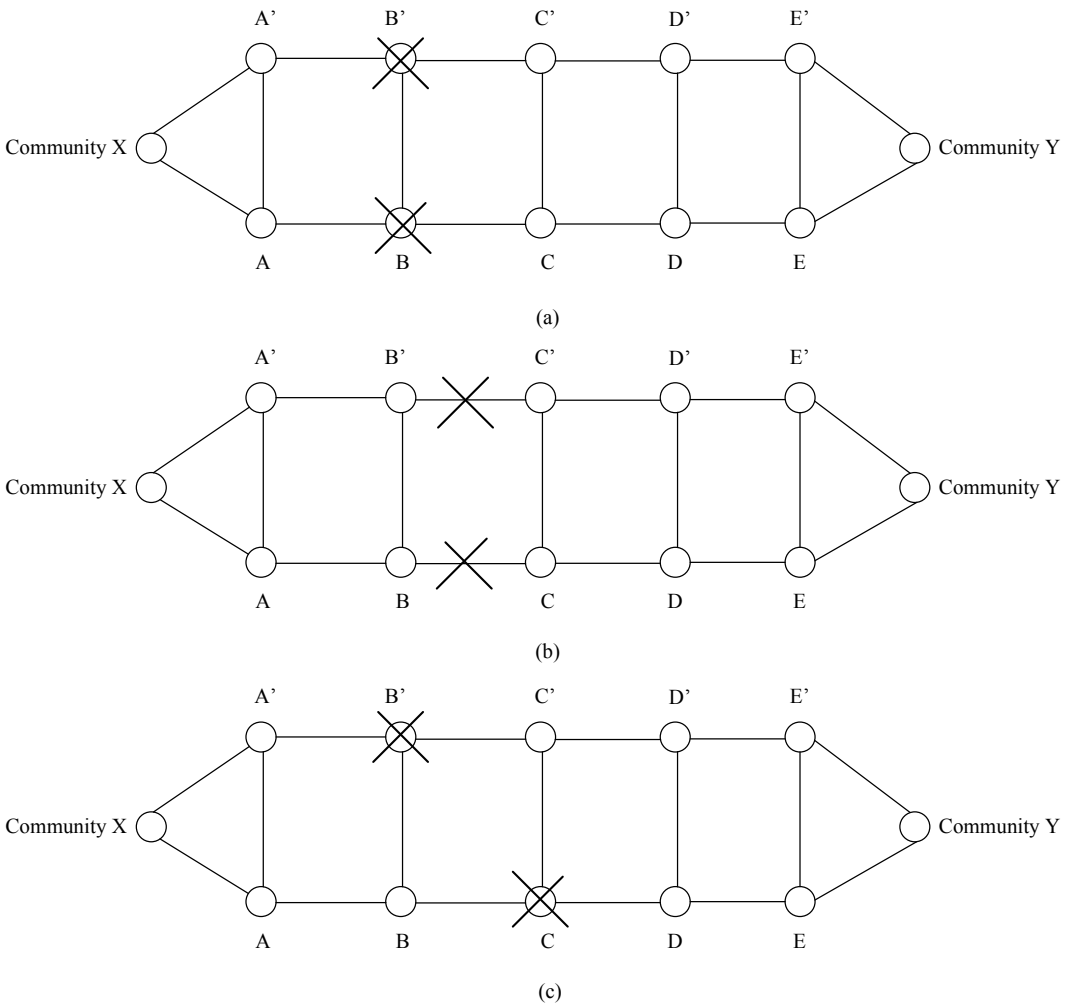



Figure 4-10 The ladder topology will not be able to overcome three failure scenarios: (a) concurrent failures of the nodes on a cross links; (b) simultaneous link failures occurring at the same hop level in the two branches; (c) failures of two nodes in consecutive hop level across the two branches.

In addition to providing an acceptable level of failure resilience, it is also essential that any wireless backhaul should be able to deliver good QoS, in terms of high throughput and low transmission delay. In the following section, the performance of a wireless backhaul network, based on the ladder topology, is evaluated by means of computer simulations.

## 4.4 PERFORMANCE EVALUATION OF THE PROPOSED LADDER TOPOLOGY

### 4.4.1 Simulation Settings

The two-branch ladder topology, as shown in Figure 4-8, is used to implement a multi-hop failure resilient wireless backhaul operating in the coordinated distributed scheduling mode as specified in IEEE 802.16 standard. Its performance is evaluated using the NCTUns network simulator [95]. First, performance evaluation is carried out when the backhaul network is operating normally with different hop counts ranging from two to eight. Constant bit rate (CBR) User Datagram Protocol (UDP) traffic is used as the data source at the gateway node, X. From node X, the traffic is divided equally and delivered to nodes A and A' to travel along the two branches of the ladder network to finally arrive at the destination node, Y.

The NCTUns network simulator employs a simple procedure for a node to search for available minislots starting from the first minislot in a frame. If there is no available minislot in the frame, the node will repeat the search in the next frame again starting from the first minislot. This process continues until either a set of available minislots required by the node has been identified, or all the minislots are checked. Furthermore, a shortest path routing protocol is adopted in the simulator to direct traffic to travel along the shortest paths available to minimise the transmission delay. For the ladder topology, the most direct or shortest paths for traffic to travel from node X to node Y are formed by the intermediate relay nodes A', B', C', D', E' and A, B, C, D, E, along the two branches of the network. In addition, several parameters need to be set in order to carry out the simulation and they are described in the following sections.

#### 4.4.1.1 Number of control transmission opportunities

As a node will have to compete with its two-hop neighbours in order to obtain a transmission opportunity for sending its control message, it is therefore necessary to provide a sufficient number of transmission opportunities to allow every node in the network to be able to transmit a control message in each frame. By examining the ladder topology with different hop counts up to eight, it is observed that a node could

have a maximum number of eight two-hop neighbouring nodes. As such, the parameter, MSH-CTRL-LEN, which represents the number of control transmission opportunities required in the frame structure for operating in the IEEE 802.16 mesh mode, is specified as eight. Since all the transmission opportunities are used to transmit Mesh Distributed Schedule (MSH-DSCH) messages, the parameter, MSH-DSCH-NUM, which corresponds to the number of transmission opportunities used to transmit the MSH-DSCH messages, is also set at eight.

#### 4.4.1.2 Reservation frame length

It is expected that a wireless backhaul network will most of the time carry a large amount of traffic. Under this condition, traffic will continuously have to be relayed by each of the network nodes. It is therefore reasonable for each of these nodes to request the maximum allowable number of frames during a three-way (TW) handshake to allow it to transmit its traffic continuously. According to the IEEE 802.16 standard, as discussed in Section 3.3, the maximum number of frames that a node can request is 128. Hence, this value of 128 is adopted as the reservation frame length for the simulation.

#### 4.4.1.3 Frame duration

A compromise value of 10 ms is chosen for the frame duration. The use of a longer frame duration is likely to increase transmission delay as a node would have to wait for a longer period before it can transmit in the next frame. Also, a shorter frame length means that there are less minislots available for data transmission [62]. The number of minislots contained in a frame is governed by the frame duration and MSH-CTRL-LEN, such that

$$\text{No. of minislots per frame} = \frac{\text{OFDM symbols per frame} - \text{MSH-CTRL-LEN} \times 7}{\left\lceil \frac{\text{OFDM symbols per frame} - \text{MSH-CTRL-LEN} \times 7}{256} \right\rceil} \quad \ldots (4.1)$$

where $\lceil \bullet \rceil$ stands for rounding up to the nearest integer. The number of OFDM symbols per frame is equal to $\text{floor}(T_F/T_s)$ with $T_F$ being the frame duration and $T_s$ is the symbol duration. In the NCTUns simulator, $T_s$ = 13.89 $\mu s$. From equation (4.1), the total number of minislots per frame becomes

$$\text{No. of minislots per frame} = \frac{\text{floor}\left(10\times10^{-3}/13.89\times10^{-6}\right)-(8\times7)}{\left\lceil\frac{\text{floor}\left(10\times10^{-3}/13.89\times10^{-6}\right)-(8\times7)}{256}\right\rceil}$$
$$= 221$$

#### 4.4.1.4 Data packet size and buffer size

A large packet size of 1000 bytes is adopted to allow us to determine the maximum throughput that could be supported by the proposed multi-hop wireless backhaul network. In addition, a sufficiently large buffer queue length is needed to reduce the possibility of buffer overflows, thereby giving rise to packet losses. For the simulation, a buffer queue length of 1000 packets is used. A smaller buffer queue length will result in frequent buffer overflows even under small traffic load conditions.

#### 4.4.1.5 Number of minislots for each link

The maximum number of minislots that could be allocated to a transmission link varies with the number of hops in the network. For fairness, individual links will have to equally share the 221 minislots available per frame. To determine the actual request size for a given link, it is necessary to first derive the collision domain set (CDS) for each link in the network. Here, the CDS of a particular link is defined as the number of links, including itself that is potentially in conflict for channel resources. For example, in a four-hop ladder topology, as shown in Figure 4-11, link $L_3$ is associated with the largest CDS of 7. Within this CDS, there are links which are allowed to transmit concurrently without giving rise to collisions, and these links are highlighted with the same colour in Figure 4-11. Now, if all the links are to be allowed to transmit simultaneously, then only those same coloured links are allowed to make use of the same set of minislots in a frame. When this is taken into account,

the largest CDS for the 4-hop network of Figure 4-11 will be reduced from seven to four. This means that only four distinct sets of minislots may be available to allow individual links to transmit data without causing collisions.

It follows that the maximum request size for a link in this four-hop ladder network is given by

$$\text{Request size} = \left\lfloor \frac{\text{Total number of data minislots in a frame}}{\text{Number of required minislots sets in the modified largest CDS}} \right\rfloor$$
$$= \left\lfloor \frac{221}{4} \right\rfloor = 55 \text{ minislots} \qquad \ldots (4.2)$$

where $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer.

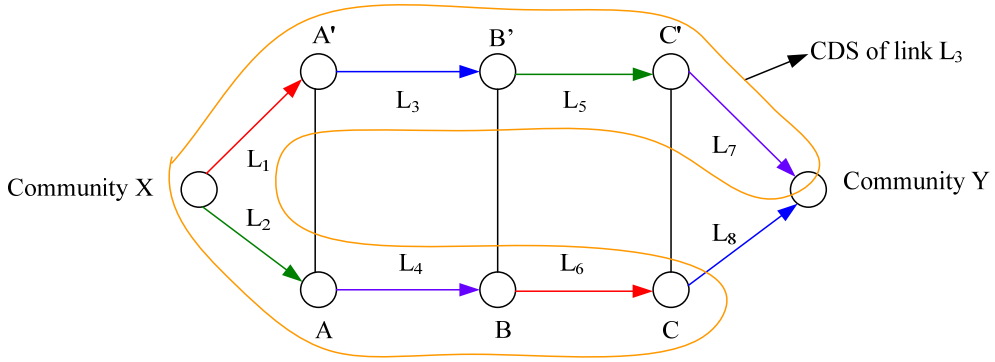



Figure 4-11 The largest CDS in a four-hop ladder topology is observed at link $L_3$ with the value of 7. Within this CDS, the links, which can transmit data simultaneously in the four-hop ladder topology, are highlighted with the same colour. Note that the cross links are not used under the shortest path routing.

Figure 4-12 shows the distribution of these four sets of minislots for data transmission by each of the links based on the above calculated request size. This particular example, however, also represents the best scenario of distributing minislots in a four-hop ladder network such that each link is being served by the maximum allowed number of minislots.

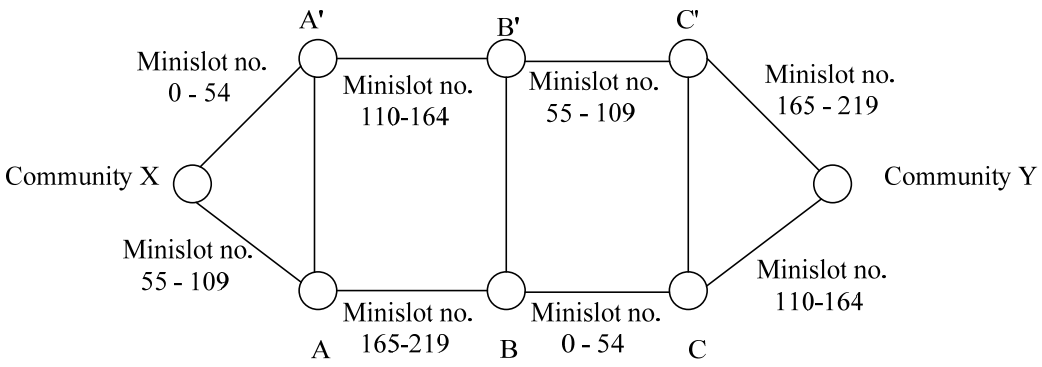



Figure 4-12 The calculated request size allows each link to obtain 55 minislots.

Now, as a given node is required to coordinate its data transmission schedule with its one-hop and two-hop neighbours through the IEEE 802.16 three-way handshake, there is a possibility that it may occasionally fail to get any share of the minislots, if all the minislots have already been taken up by its neighbours. One such situation is typified by the particular pattern of distribution of the minislot sets in the four-hop ladder topology of Figure 4-13(a). In this case, node B' has to refrain from transmission, and eventually its accumulated data packets will overflow its buffer. In order to avoid this type of situation, it becomes necessary to increase the number of minislot sets to be used for distribution among the individual links. Moreover, this will also lead to a reduction in the allowable request size with the consequence of possible lowering of the maximum achievable throughput for the backhaul network.

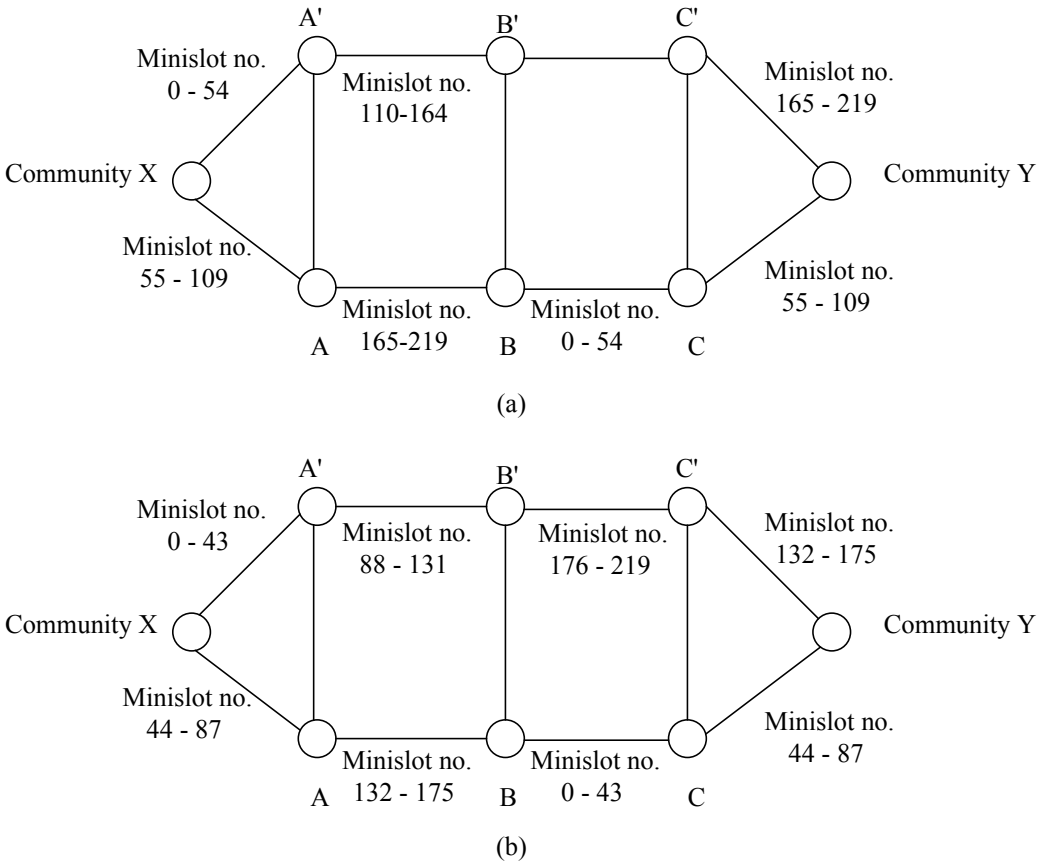



Figure 4-13 Minislot allocation in a four-hop ladder topology: (a) node B' fails to get any minislots with the calculated request size of 55 as all the minislots are used by the other links; and (b) the request size is reduced to 44 to allow every link to obtain minislots.

As the number of hops in the network increases, the likelihood of the above undesirable situation, illustrated in Figure 4-13(a), occurring is also increased. However, actual occurrences of such an event are difficult to predict. As such, it is proposed that the appropriate request size for use in a ladder network with a given number of hops is derived based on the following procedure:

- Use equation (4.2) to first calculate the maximum request size for a given hop count.
- Based on the above calculated request size, verify its suitability with computer simulations. If the problem exists, increase the number of minislot sets by one, and recalculate the request size. Repeat this step until the problem is not observed.

The resultant request size obtained in this way will not only substantially reduce the problem associated with an event of Figure 4-13(a) but also ensure that its value will result in a throughput which is as large as possible. Table 4-1 tabulates the request sizes obtained in this way for ladder networks of different hop counts.

Table 4-1The request size for a given number of hops.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 55 |
| 3 | 55 |
| 4 | 44 |
| 5 | 44 |
| 6 | 36 |
| 7 | 36 |
| 8 | 36 |

#### 4.4.1.6 Traffic data bit rate

Constant bit rate (CBR) User Datagram Protocol (UDP) traffic is used as the data source for performance evaluation. Unlike the TCP-IP traffic model, its generation does not involve any specific handshaking procedures. As such, the results from the performance evaluation are influenced only by the IEEE 802.16 three-way handshake, and not constrained by the way traffic is generated. As discussed earlier in this section, the actual request size that could be used by a given node in the ladder network varies with the number of hops between the source and destination gateway nodes. This means that the maximum bit rate of the UDP traffic that could be supported by the proposed ladder network of a given hop count is related to the request size used, such that

$$CBR = \frac{\text{No. of bits transmitted in a frame, } N_B}{\text{Frame duration, } T_f}. \tag{4.3}$$

where $N_B = r_s \times m_b - o_h$ with $r_s$ being the request size, $m_b$ the number of bits that can be transmitted in a minislot, and $o_h$ the overhead.

The CBR value computed from equation (4.3) corresponds to the maximum traffic load that could be supported by a given network before packet drops start to occur due to buffer overflow. In practice, it is expected to realise a lower CBR value. In this study, the actual CBR value adopted for use in the proposed ladder network of a given hop count has been obtained by gradually reducing its value, starting from the computed CBR value, until a very small packet loss of not more than 0.003 % begins to appear.

#### 4.4.1.7 Parameters summary

The various parameters adopted for the performance evaluation by computer simulation using the NCTUns network simulator are tabulated in Table 4-2.

Table 4-2 Simulation parameters.

| Parameter | Value |
|---|---|
| MSH-CTRL-LEN | 8 |
| MSH-DSCH-NUM | 8 |
| Reservation frame length | 128 |
| Frame duration | 10 ms |
| Number of minislots per frame | 221 |
| Total number of packets | 600000 |
| Number of runs | 10 |
| Packet size | 1000 bytes |
| Queue buffer length | 1000 packets |

### 4.4.2 Simulation Results

Two performance metrics are used to evaluate the performance of the proposed wireless backhaul network, namely maximum achievable throughput, and average end-to-end packet transmission delay. The former is determined based on the

maximum traffic load that the network is able to support, while maintaining no or near zero packet loss. Also, the average end-to-end packet transmission delay is measured when the maximum achievable throughput is reached. For reliable performance measurements, both maximum achievable throughput and average end-to-end packet transmission delay are given as the ensemble averaged values of ten simulation runs, each consisting of 600 Mbytes.

Based on the request size, as tabulated in Table 4-1, the theoretical maximum traffic load, $Tload_t$, that can be applied to a ladder backhaul network of a given hop count is given by

$$Tload_t = \frac{2 \times r_s \times k \times m}{T_F} \tag{4.4}$$

where $r_s$ is the request size, $T_F$ is the frame duration, and $m$ is the number of bits that can be transmitted in an OFDM symbol. For 64 QAM-3/4 modulation scheme, $m$ is equal to 824 bits [35]. The value of $k$ is obtained as

$$k = \left\lceil \frac{\text{Number of OFDM symbols in a frame - MSH\_CTRL\_LEN} \times 7}{256} \right\rceil \tag{4.5}$$

where $\lceil \bullet \rceil$ stands for rounding up to the nearest integer. Table 4-3 shows the resultant theoretical maximum traffic loads that can be supported by ladder backhaul networks of two to eight hops. As expected, the maximum traffic load that can be supported decreases when the hop count is increased. This is due to the fact that a ladder network of a larger hop count has to make use of a smaller request size.

Table 4-3 Maximum traffic loads that can be supported by ladder backhaul networks of different hop counts.

| Number of hops | Request size (Minislots) | Maximum traffic load (Mbps) |
|---|---|---|
| 2 | 55 | 27.19 |
| 3 | 55 | 27.19 |
| 4 | 44 | 21.75 |
| 5 | 44 | 21.75 |
| 6 | 36 | 17.80 |
| 7 | 36 | 17.80 |
| 8 | 36 | 17.80 |

The IEEE 802.16 TW handshake procedure is applied for coordinated distributed scheduling in a wireless backhaul based on the proposed ladder topology. Using the parameters of Table 4-2, the performance of the wireless backhaul network implemented is evaluated by computer simulation. Table 4-4 shows the maximum achievable throughputs and average end-to-end packet transmission delays achieved with different hop counts, ranging from two to eight hops.

Table 4-4 Maximum achievable throughput and averange end-to-end packet transmission delay when the IEEE 802.16 TW handshake is used in the proposed ladder network of different hop counts.

| Number of hops | Request size | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 55 | 25.80 | 70.23 |
| 3 | | 11.29 | 24.16 |
| 4 | 44 | 0.82 | 317.73 |
| 5 | | 0.86 | 365.75 |
| 6 | 36 | 0.84 | 384.04 |
| 7 | | 0.81 | 515.82 |
| 8 | | 0.62 | 540.66 |

As expected the two-hop backhaul can achieve the best maximum achievable throughput. However, the network throughput is rapidly degraded as the number of hops increases. For example, the throughput degrades by 92.7 % when the hop count increases from three to four. In this case, there is a sudden jump in the number of occurrences from three to seven for cases involving two or more hidden nodes when the hop count is increased from three to four, as shown in Table 4-5. Such a situation will reduce the number of minislots allocated to each individual node, leading to a decrease in the maximum achievable throughput. Moreover, as the number of hops is increased each time beyond four, the increase in the number of occurrences of 2 or more hidden nodes tends to be less drastic. As such, the achievable throughputs with these networks remain similar as observed from Table 4-4.

Table 4-5 Number of occurences of hidden nodes associated with a multi-hop ladder backhaul network for hop counts of two to six.

| Number of hidden nodes | Number of occurrences | | | | |
|---|---|---|---|---|---|
| | Two-hop | Three-hop | Four-hop | Five-hop | Six-hop |
| 1 | 0 | 2 | 0 | 0 | 0 |
| 2 or above | 0 | 3 | 7 | 9 | 11 |

The result tabulated in Table 4-4 clearly suggests that the IEEE 802.16 TW handshake process is ineffective for application in a multi-hop network. For a ladder network with more than two hops, the hidden node problem will impact the TW handshaking procedure. As a result, the bandwidth requested by a given node may either not be granted or under granted.

Data packets, which have to traverse a larger number of hops before arriving at the destination, tend to experience higher average packet transmission delays, as indicated in Table 4-4. However, comparison of transmission delays for different hop counts becomes more complicated due to the different request sizes used. Strictly, a fair comparison can only be made when networks are subjected to a similar traffic load, i.e., adopting the same request size. Moreover, simulation shows a rather unexpected outcome that the average transmission delay associated with the three hop network is much lower than that for the two hop network. This can be explained on closer examination of the throughputs achieved with these two networks. Since

the throughput of the two hop network is more than twice that of the three hop network, this means that the intermediate relay nodes of the former are scheduling twice as much traffic load as those of the latter. With this increase in traffic, it is likely that the average buffer queue of the two hop network will increase, leading to a longer average end-to-end delay. In fact, it is observed from the simulation that the average buffer queue lengths of the two and three hop networks are 24 and 6 packets, respectively. Also, as shown in Table 4-5, as the number of hops in a ladder network increases, the hidden node problem becomes even more prominent. Consequently, the hidden node problem will reduce the effectiveness of the IEEE 802.16 TW handshake procedure in allocating minislots to individual nodes of the network. As such, this leads to a large drop in throughput accompanied by a drastic increase of average end-to-end transmission delay. It is clear from the results of Table 4-4 that the existing IEEE 802.16 TW handshake process is not useable for a ladder network greater than three hops. Proposed improvements to the IEEE 802.16 TW handshake process are presented in the later parts of this chapter, and the next chapter.

### 4.4.3 Hidden Node Problem Associated with the IEEE 802.16 Three-way Handshake Protocol

As discussed in Section 3.4.3, the IEEE 802.16 three-way handshake protocol suffers from a serious hidden node problem when applied to a multi-hop ladder network. This problem occurs when two receiving nodes, which are outside the receiving range of one another, happen to select the same minislots to be granted to their respective transmitting nodes [80, 81]. In other words, the effect of such a problem is measured in terms of the packet-reception distance rather than the signal interference distance. In practice, the latter is usually longer than the former. For example, in previous studies concerning IEEE 802.11 networks [59], the interference distance is usually set to be twice the packet-reception distance. This suggests that even if a node, say node A, cannot receive a packet sent by its immediate neighbouring node, such as node B, the signal from node B', which is two node distance away, may still affect the reception of packets at node A. Moreover, in studies concerning the IEEE 802.16 standard, it is common to consider radio frequency interference arising from

the immediate adjacent neighbouring nodes and ignore any accumulative interference from other more remote nodes [35, 96, 97].

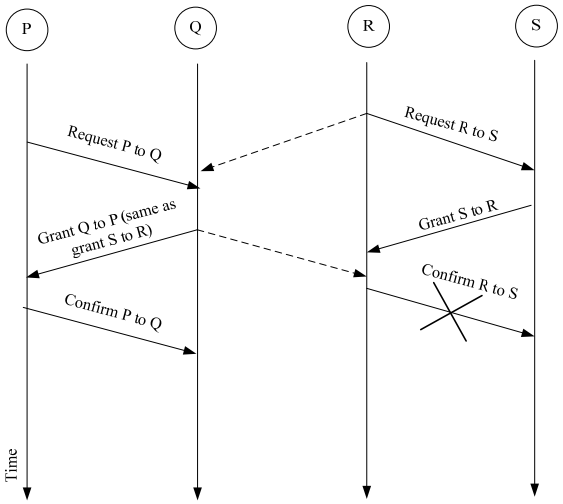



Figure 4-14 Grant withdrawal occurs when node R detects a minislot allocation conflict, which refrains it from confirming the grant from node S.

Hidden node problem will cause grant withdrawal, as illustrated in the example of Figure 4-14. In this case, both nodes Q and S select the same minislots to be granted to nodes P and R, respectively. However, node R will detect a minislot allocation conflict when it overhears the grant from node Q. Consequently, node R will not be able to confirm the grant of resource by node S. Otherwise, nodes P and R will transmit simultaneously, thus giving rise to packet collisions at node Q. Under this situation, node R will refrain from data transmission, and this will result in a reduction of overall network throughput, and an increase of end-to-end transmission delay. Such a problem may be overcome through the use of a regranting scheme [80, 81]. As shown in Figure 4-15, the use of the regranting scheme allows node S to grant a different set of minislots to node R when it does not receive a confirm IE from node R. By doing so, node R will be able to transmit data after confirming the new grant. Nonetheless, as discussed in Section 3.4.3, grant withdrawal only occurs if nodes do not transmit their respective availability IEs when they make requests for bandwidth. By transmitting an availability IE during a TW handshake, a node can prevent its neighbouring nodes from requesting or granting the same minislots.

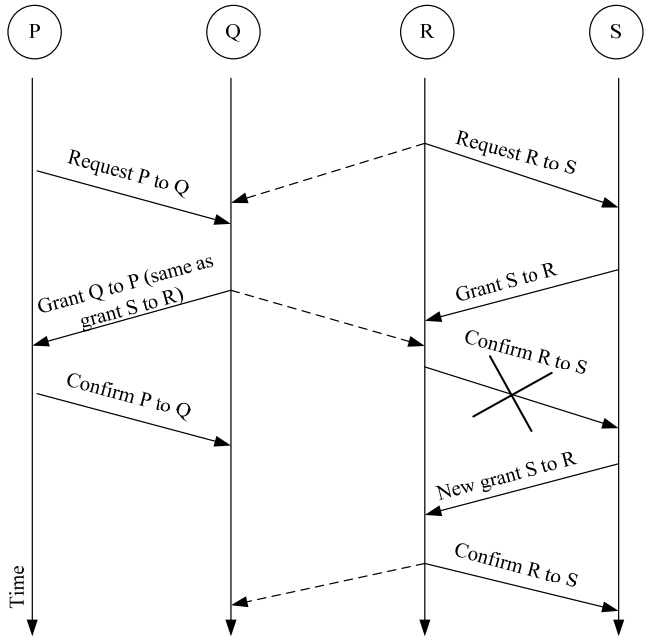



Figure 4-15 Regranting scheme.

On the other hand, the transmission of an availability IE during a TW handshake will not necessarily eliminate the hidden node problem. One such scenario is shown in Figure 4-16. In this example, node P is not aware of node R also requesting the same set of minislots. Upon receiving the request and availability IEs of node P, and overhearing those from node R, node Q will detect a minislot request conflict. Under this condition, node Q will either reject the request of node P, or grant it the requested minislots only after node R has finished its transmission to avoid packet collisions at node Q. In either case, node P is likely to wait for a long period of time before it can transmit data. Moreover, it is interesting to note that the hidden node problem does not result in grant withdrawal under the scenario depicted in Figure 4-16. Hence, the regranting scheme proposed in [80, 81] is not applicable in the situation described in this example. It is also observed that with the example of Figure 4-16, the hidden node problem involves two transmitting nodes that are separated by two hops, i.e., nodes P and node R. As the number of hops in a network increases, this will also lead to a larger number of nodes which have neighbours located two hops away. This suggests that an effective solution to the hidden node problem will need to be found for efficient data transmission in a multi-hop ladder network.

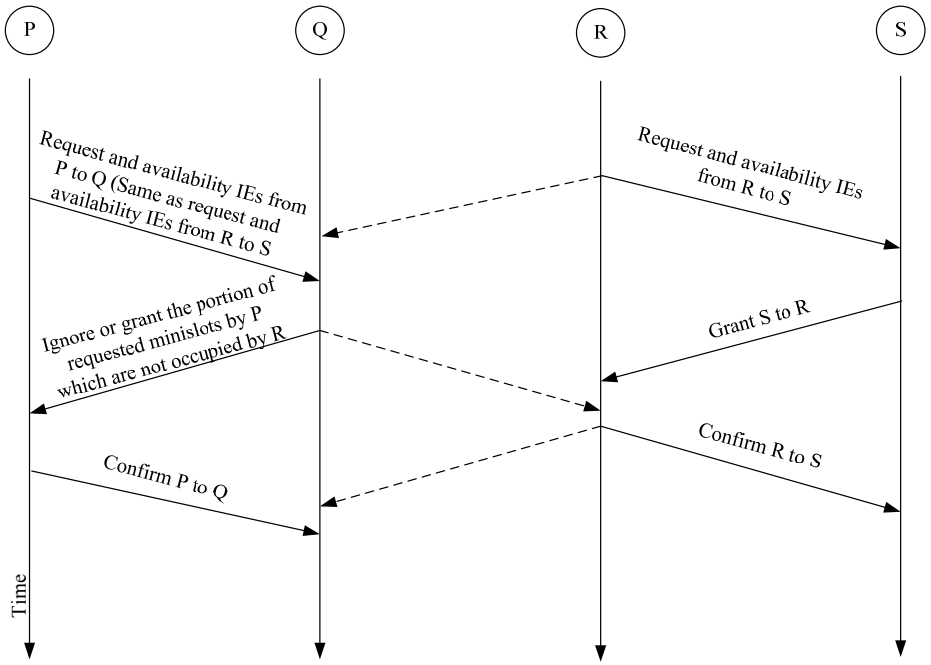



Figure 4-16 Hidden node scenario when the availability IE is used during a three-way handshake.

### 4.4.4 Proposed Reverse Notification Control Message

In this section, an extension is made to the IEEE 802.16 TW handshaking protocol in an attempt to overcome the hidden node problem associated with multi-hop transmission. This involves the use of a new reverse notification (RN) control message, which takes the form of a duplicate copy of an availability IE. With this proposed scheme, a RN control message is sent only after a node has received request and availability IEs that are not destined for it. This control message can be sent in conjunction with either one of the three IEs, i.e., request, grant or confirm, within the control subframe. This is made possible by the fact that a node is able to send multiple IEs via a single MSH-DSCH message at a given time [35]. In this way, it significantly increases the likelihood for the destined node to receive the RN message before it attempts to ask for channel resources.

An example of the exchange of a RN message is shown in Figure 4-17. It shows that node Q, which overhears the transmission of request IE by node R, will transmit a RN to node P. Upon receiving this message, node P will request only those minislots which are not listed in the RN IE. By doing so, node P will be able to obtain its requested minislots from node Q. For the hidden node problem to be handled

effectively, all the neighbouring nodes of node R are also required to send RN messages. As the information capacity required in a RN message, which is piggybacked on other IEs, is small, the extra overhead and delay incurred are negligible.

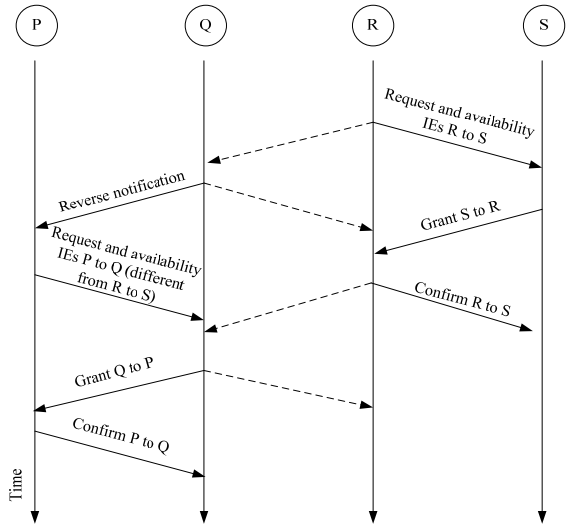



Figure 4-17 An exchange of RN control message.

Nevertheless, the inclusion of the RN control message in the TW handshake still does not completely resolve the hidden node problem. For example, if node P fails to receive a RN message in time before it makes a request for channel resources, then it is possible that it will request the same minislots that have already been used by its neighbouring nodes. This situation is illustrated in Figure 4-18. However, it is observed from computer simulations that the number of occurrences of such an event is rather infrequent when the network is operating under normal condition. As such, it is expected that the proposed extension to the TW handshake will be able to significantly enhance the network throughput.

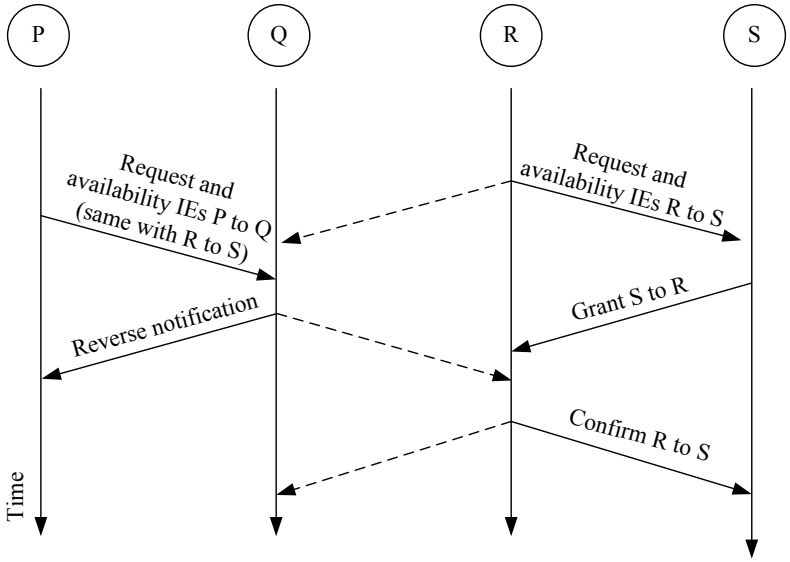



Figure 4-18 A scenario where the RN scheme fails to prevent two two-hop neighbouring nodes from making the same resource request.

Due to the broadcast nature of a wireless environment, other nodes in the proximity of the one that broadcasts the RN message, say node Q in this case, might also receive the same control message. In such a situation, those nodes that intend to send data to node Q will either have to defer making requests for minislots, which have already been occupied, or request some other available minislots. This is to avoid possible collisions at node Q. On the other hand, those nodes that do not intend to send a request to node Q will simply ignore the RN control message.

### 4.4.5 Performance Evaluation of the Reverse Notification Scheme

The performance of the proposed RN scheme operating in a ladder topology has been evaluated using the same simulation settings as given in Section 4.4.1. Figure 4-19 and Table 4-6 show the maximum achievable network throughputs obtained via the modified and the original IEEE 802.16 coordinated distributed scheduling schemes.

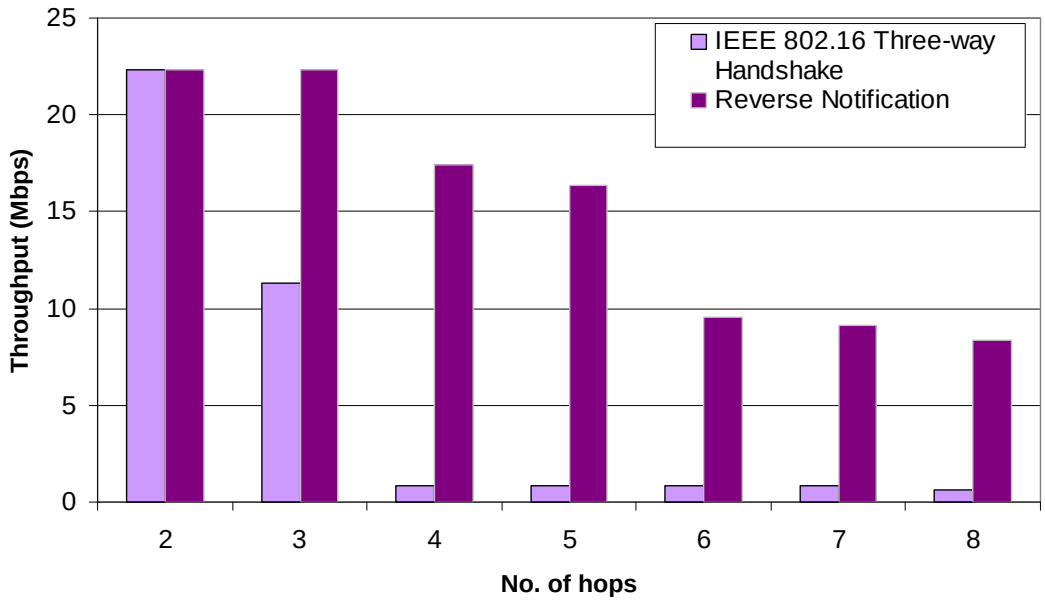



Figure 4-19 Maximum achievable throughputs obtained with different hop counts.

Table 4-6 The maximum achievable throughput for the ladder topology with different hop counts operating under IEEEE 802.16 three-way handshake and RN.

| Number of hops | Request size (minislots) | Throughput (Mbps) | |
|---|---|---|---|
| | | IEEE 802.16 TW handshake | RN |
| 2 | 55 | 25.80 | 25.80 |
| 3 | | 11.29 | 25.80 |
| 4 | 44 | 0.82 | 17.45 |
| 5 | | 0.86 | 16.38 |
| 6 | 36 | 0.84 | 9.55 |
| 7 | | 0.81 | 9.11 |
| 8 | | 0.62 | 8.36 |

As discussed in Section 4.4.3, the hidden node problem occurs among sending nodes, which are located two hops from one another. Now, in the special case of the two-hop ladder network, as shown in Figure 4-20, all the sending nodes, i.e., Community X, node A, and node A', are only one hop away from each other. In this case, the nodes will not encounter the hidden node problem. Therefore, it is expected that the same throughput should be achieved with the two-hop ladder network irrespective of

whether the original or enhanced TW handshaking process is used. Computer simulation results of Figure 4-19 and Table 4-6 have indeed verified this observation.

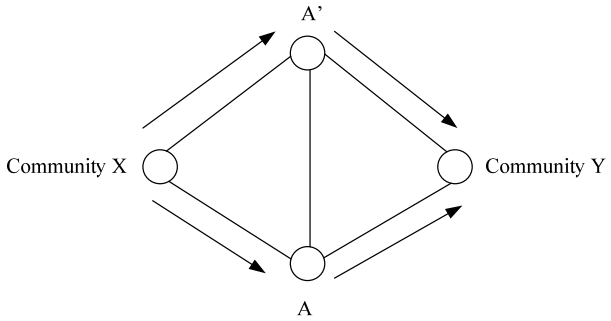



Figure 4-20 As all the sending nodes in the two-hop ladder topology, X, A, and A', are within one-hop of each other, they will not encounter the hidden node problem.

For ladder networks with more than two hops, the effectiveness of the proposed RN scheme in mitigating the hidden node problem is well demonstrated by the greatly improved throughputs achieved, as shown in Figure 4-19. For instance, the use of RN has enabled the three-hop network to achieve the same maximum achievable throughput as its two-hop counterpart, thus indicating that the hidden node problem is, for practical purposes, mitigated. This observation clearly suggests that the hidden node problem is a major cause to the significant throughput degradation in multi-hop ladder networks when the IEEE 802.16 TW handshake procedure is used for traffic scheduling. As the number of hops is further increased beyond three, the number of nodes, which have neighbours located two hops away, also becomes larger. This will not only aggravate the hidden node problem but also reduce the request size of individual nodes. The former is due to the large contention for transmission opportunity experienced by a node when the number of neighbouring nodes is increased. A consequence of this is that less RN IEs will be sent. This then lowers the likelihood of the RN IEs being received by the relevant nodes before they send their request. Consequently, it reduces the effectiveness of the RN scheme in mitigating the hidden node problem. The combined effects of the hidden node problem and smaller request size result in a reduction in the maximum throughput that could be achieved for a ladder network with a larger hop count. In fact, the results tabulated in Table 4-3 and Table 4-6 show that the maximum achievable throughputs of the wireless backhaul with hop counts larger than three are still far

from the theoretical maximum traffic loads that can be supported. Nonetheless, the maximum achievable throughput for an 8-hop ladder network achieved through the use of RN is almost 14 times larger than that achieved using the original IEEE 802.16 three-way handshake.

Table 4-7 Average end-to-end transmission delay for different ladder topologies.

| Average end-to-end delay (ms) | | | |
|---|---|---|---|
| Number of hops | Request Size (minislots) | IEEE 802.16 TW handshake | RN |
| 2 | 55 | 70.23 | 69.67 |
| 3 | | 24.16 | 94.15 |
| 4 | 44 | 317.73 | 104.78 |
| 5 | | 365.75 | 83.45 |
| 6 | 36 | 384.04 | 53.71 |
| 7 | | 515.82 | 60.18 |
| 8 | | 540.66 | 74.77 |

In addition to improving throughputs, the use of the proposed RN scheme also helps to significantly lower the average end-to-end packet transmission delay of multi-hop ladder networks. For example, as shown in Table 4-7, it is interesting to observe that the average end-to-end packet transmission delays achieved with RN for all the networks considered are less than 150 ms, which is usually considered as the upper limit allowed for services involving delay sensitive real-time traffic. Apparently, there is an increase in delay when the enhanced TW handshaking procedure, instead of the original one, is used in the 3-hop network. This is largely due to the almost four times increase in achieved throughput or equivalent traffic load handled by the enhanced TW handshaking scheme. As a result, packets are dropped in the three-hop network due mainly to buffer overflow. In other words, an increasing larger number of packets are being stored in the buffers of individual nodes, giving rise to queueing delay which forms part of the average end-to-end transmission delay. In general, a network with a larger hop count is expected to experience an increased average end-to-end transmission delay, as the traffic has to traverse a longer path before reaching

the final destination. However, it is surprising to note that the average end-to-end packet transmission delays for hop counts beyond four are lower than that of the four-hop ladder network. For those networks, the aggravated hidden node problem, due to the increased number of neighbouring nodes, has reduced the amount of traffic load that can be applied to the networks. In particular, the traffic load applied in the six-hop ladder network to yield the maximum achievable throughput tabulated in Table 4-6 is only half of that for the five-hop ladder network. This suggests that fewer packets are in the buffer queue of each individual node in the six-hop ladder network, and the packets will experience a smaller transmission delay.

Overall, the proposed RN scheme is shown to be very effective in combating the hidden node problem, thereby greatly enhancing the maximum achievable throughputs, and reducing the average end-to-end packet transmission delays of multi-hop ladder networks.

## 4.5 SUMMARY

In this chapter, a failure resilient multi-hop IEEE 802.16 wireless backhaul is proposed to interconnect a distant remote community to a gateway node located in the regional or metropolitan centre. This wireless backhaul network involves long routes with few users surrounding the intermediate nodes. Several design criteria are identified for determining the feasible node locations as well as the interconnection links between these nodes. These criteria include network cost, failure scenarios, level of connectivity, interference, traffic rerouting strategy, and transmission delay incurred during traffic rerouting. The proposed wireless backhaul takes the form of a ladder network topology, which provides at least one backup path for each node pair. This ladder topology is able to deliver the coverage between the two community centres by making use of a minimum number of nodes while still providing the necessary backup paths. With the exception of a few failure scenarios, the proposed wireless backhaul network is able to sustain multiple link and node failures.

Moreover, the performance of the ladder network operating with the original IEEE 802.16 three-way handshake protocol is evaluated, by means of computer

simulations using the NCTUns network simulator, in terms of the maximum achievable throughput and average end-to-end transmission delay. This has been carried out for ladder networks with hop counts ranging from two to eight. The results are tabulated in Table 4-4, which shows that the original TW handshaking protocol suffers from the hidden node problem, which is well known in multi-hop wireless networks. As a result, a four-hop ladder network can only manage to achieve a maximum throughput of 0.82 Mbps. At the same time, data packets experience a long average end-to-end transmission delay in excess of 300 ms. The results of Table 4-4 also provide the necessary reference for comparison with other proposed schemes described in this thesis.

A new reverse notification (RN) scheme, which proves to be very effective in mitigating the hidden node problem, is proposed in this chapter. Computer simulations have verified that great improvement in maximum achievable throughputs, as well as reduction in average end-to-end transmission delays, are able to be achieved with this new RN scheme. The computer simulated results are shown in Figure 4-19 and Table 4-7. Again, for a four-hop ladder network, the use of the RN scheme has increased the maximum achievable throughput by almost 22 times while reducing the average end-to-end transmission delay by 69%, when compared with the use of the original TW handshake protocol. The results also suggest that the ladder topology with different hop counts, when incorporated with the proposed enhanced IEEE 802.16 TW handshake protocol, is suitable for realising a broadband wireless backhaul network for delivering services involving delay sensitive real-time traffics.

So far, the results presented in this chapter have been obtained under the condition that the hidden node problem is measured in terms of packet-reception distance. In the case when the radio interference distance is taken into consideration, the performance of RN is likely to be degraded by 20% [96]. Such a relatively small degradation in performance will not overshadow the significant improvement in network throughput achieved using the proposed RN. Moreover, the results are obtained when the network is operating without any node or link failure. In the next chapter, the performance of the proposed RN scheme operating under failure conditions will be investigated.

# CHAPTER 5

# PERFORMANCE OF AN IEEE 802.16 WIRELESS BACKHAUL IN THE PRESENCE OF FAILURE

## 5.1 INTRODUCTION

It is shown in Chapter 4 that the serious hidden node problem associated with the IEEE 802.16 three-way (TW) handshake protocol operating in a multi-hop ladder network has largely been overcome by incorporating the proposed reverse notification (RN) scheme within the TW handshake process. As a result, the performance of the network, in terms of throughput and end-to-end transmission delay, is significantly enhanced. In this chapter, the performance of the ladder network, operating as a wireless backhaul, is evaluated in the presence of failures in either link or node.

This chapter is organised as follows. First, the operation of the ladder network employing the IEEE 802.16 standard coordinated distributed scheduling, in conjunction with the proposed RN scheme, is investigated in Section 5.2. Further new modifications to the TW handshake process are presented in Section 5.3 to overcome some problematic scenarios which are likely to affect the performance of the wireless backhaul in the event of a network failure. Computer simulated results achieved through the use of these new schemes under normal operation and failure conditions are presented in Section 5.4. Also, the effects on the network performance due to the use of different buffer and packet sizes are examined. The performance of the ladder network is then compared with the two parallel path network in Section 5.5. So far, traffic is considered to be flowing in one direction, i.e., from Community X to Community Y. In practice, a wireless backhaul has to be able to handle traffics emanating from both directions. As such, the performance of the ladder network involving bidirectional traffic flows is also evaluated in Section 5.6.

## 5.2 OPERATION OF THE LADDER NETWORK IN THE EVENT OF A NODE FAILURE

Consider a four-hop ladder network as shown in Figure 5-1(a). Under normal operating conditions, for node X, there are four other nodes within its 2-hop neighbourhood. However, this number is reduced to two when node A fails, as shown in Figure 5-1(b). With this reduced number of nodes within its 2-hop neighbourhood, node X is likely to experience fewer contentions for transmission opportunity from the other nodes. It then becomes possible for node X to be able to send a request for resource to its immediate neighbour, node A' in this case, before receiving a RN message from it. This scenario, as pointed out in Section 4.4.4, is one that will make the proposed RN scheme ineffective in overcoming the hidden node problem. The above is also true if the failure occurs at node A instead of A'.

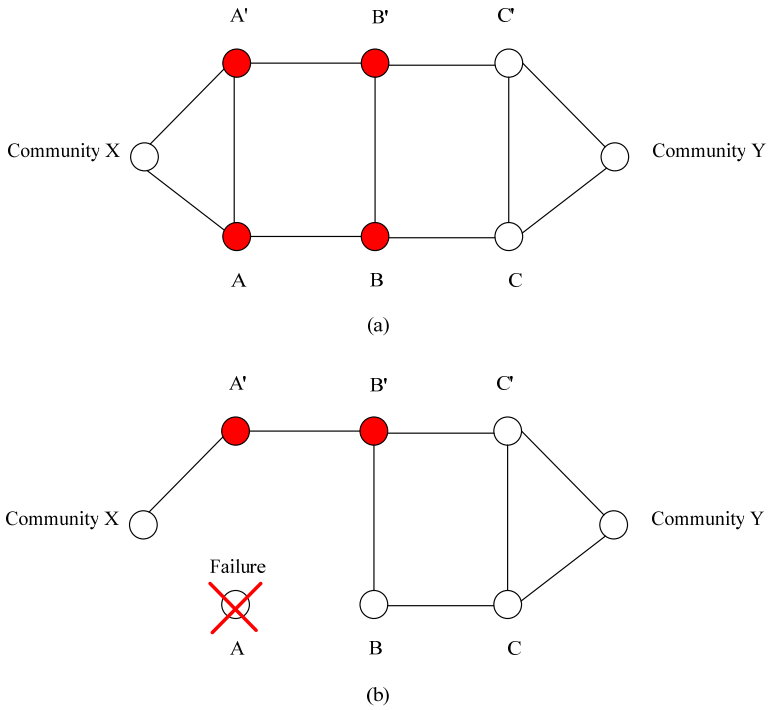



Figure 5-1 The nodes, coloured in red, are within two-hop away from node X, during: (a) normal operating condition; (b) when node A fails.

On the other hand, if any one of the other intermediate nodes (B, B', C, and C') of the 4-hop ladder network were to fail, it is possible for a given normal operating node in the network to be associated with a different number of potential hidden nodes. To verify this observation, we set out to identify all the potential hidden nodes

associated with any of the normal operating nodes when one particular node in the network becomes faulty. The results are tabulated in Tables 5-1 to 5-6.

Table 5-1 Hidden nodes encountered by a given node when node A fails.

| Node | Potential hidden nodes | Number of hidden nodes |
|---|---|---|
| X | B’ | 1 |
| A’ | B, C’ | 2 |
| B’ | X, C | 2 |
| B | A’, C’ | 2 |
| C’ | A’, B | 2 |
| C | B’ | 1 |

Table 5-2 Hidden nodes encountered by a given node when node A' fails.

| Node | Potential hidden nodes | Number of hidden nodes |
|---|---|---|
| X | B | 1 |
| A | B’, C | 2 |
| B’ | A, C | 2 |
| B | X, C’ | 2 |
| C’ | B | 1 |
| C | A, B’ | 2 |

Table 5-3 Hidden nodes experienced by a given node when node B fails.

| Node | Potential hidden nodes | Number of hidden nodes |
|---|---|---|
| X | B’ | 1 |
| A’ | C’ | 1 |
| A | B’ | 1 |
| B’ | X, A, C | 3 |
| C’ | A’ | 1 |
| C | B’ | 1 |

Table 5-4 Hidden nodes encountered by a given node when node B' fails.

| Node | Potential hidden nodes | Number of hidden nodes |
|---|---|---|
| X | B | 1 |
| A’ | B | 1 |
| A | C | 1 |
| B | X, A’, C’ | 3 |
| C’ | B | 1 |
| C | A | 1 |

Table 5-5 Hidden nodes encountered by a given node when node C fails.

| Node | Potential hidden nodes | Number of hidden nodes |
|---|---|---|
| X | B’, B | 2 |
| A’ | C’, B | 2 |
| A | B’ | 1 |
| B’ | X, A | 2 |
| B | X, A’, C’ | 3 |
| C’ | A’, B | 2 |

Table 5-6 Hidden nodes encountered by a given node when node C’ has failed

| Node | Potential hidden nodes | Number of hidden nodes |
|---|---|---|
| X | B’, B | 2 |
| A’ | B | 1 |
| A | B’, C | 2 |
| B’ | X, A, C | 3 |
| B | X, A’ | 2 |
| C | A, B’ | 2 |

Also, Table 5-7 summarises the number of potential hidden nodes and the corresponding number of occurrences according to the location of a node failure.

Table 5-7 Number of potential hidden nodes and corresponding number of occurrences associated with the failure of a specified node.

| Number of hidden nodes | Number of occurrences | | | | | |
|---|---|---|---|---|---|---|
| | Node A | Node A' | Node B | Node B' | Node C | Node C' |
| 1 | 2 | 2 | 5 | 5 | 1 | 1 |
| 2 or above | 4 | 4 | 1 | 1 | 5 | 5 |

It is observed from Table 5-7 that a failure that occurs at any one of the two nodes connecting the same cross link across the two branches of the ladder network, for example, nodes A and A', will give rise to the same number of hidden nodes. As such, it is expected that the failure of either one of those two nodes will have the same effects on the operation of the network. Furthermore, it is observed that when either node B or B' fails, the majority of the other normal operating nodes only encounter one single hidden node. On the other hand, when either node C or C' fails, the majority of the other nodes are going to have two or more hidden nodes. This suggests that for this 4-hop ladder network, the introduction of the proposed RN in the TW handshaking process will have a greater benefit when the failure occurs at either node B or B' instead of at node C or C'.

Next, the performance of the 4-hop ladder network, operating with the IEEE 802.16 coordinated distributed scheduling in conjunction with the proposed RN, is evaluated when a failure occurs in any one of the six intermediate nodes. Again, the same simulation settings as described in Section 4.4.1 are adopted for the computer simulations using the NCTUns network simulator.

Table 5-8 tabulates the throughputs and average delays obtained when there is a failure occurring at one of the intermediate nodes in the 4-hop ladder network. As expected, the performance of the ladder network varies according to where a node failure occurs. The results show that the highest throughput of 9.11 Mbps is obtained when either node B or B' fails. On the other hand, the network yields the least throughput of 6.81 Mbps when the failure occurs at either node C or C'. These confirm our earlier observations on the effectiveness of the proposed RN scheme

when dealing with a possible node failure in the 4-hop ladder network. The simulated results also verify that the same throughput is achieved when the failure occurs at any one of the two nodes sharing the same crosslink. The fact that there are hidden nodes present around a given node will prevent it from obtaining its requested minislots and refrain it from data transmission. Subsequently, this will cause its packet queue to build up until finally packets have to be dropped due to buffer overflow at the node. Therefore, in order to maintain near zero packet loss, which forms the basis for determining the maximum achievable throughput, it becomes necessary to lower the traffic load applied to the network. Now, with a lower traffic load, the achievable throughput is reduced, but the traffic will also experience less delay as fewer data packets are in the queue at each node. This explains the observation that when a node failure occurs at either node C or C', both the resultant throughput and delay are lowest among the results of Table 5-8.

Table 5-8 Throughput and delay achieved when a node failure occurred at a different location.

| Node failure | Throughput (Mbps) | Average end-to-end delay (ms) |
|---|---|---|
| A | 8.18 | 51.31 |
| A' | 8.18 | 52.01 |
| B | 9.11 | 78.87 |
| B' | 9.11 | 78.94 |
| C | 6.81 | 41.82 |
| C' | 6.81 | 41.85 |

These simulations help us to realise that if we wish to further enhance the throughput of the ladder network, it will be necessary to find a way to lessen the likelihood of occurrence of the hidden node problem. As a result, a request-resend procedure is proposed for incorporation into the RN scheme. This will be described in Section 5.3.1. Furthermore, when a node fails, its immediate neighbours will be required to handle an increased amount of rerouted traffic. Therefore, a dynamic minislot allocation scheme is presented in Section 5.3.2 to allow a node to be able to receive the number of minislots appropriate for the traffic load it is servicing at the time.

## 5.3 REQUEST-RESEND AND DYNAMIC MINISLOT ALLOCATION

### 5.3.1 Request-resend

As discussed in the last section, a node failure in a ladder network is likely to increase the likelihood of a RN message arriving at the destined node only after it has sent out its request for minislots. Such an event is in conflict with the intended operation of the proposed RN scheme, as described in Section 4.4.4. To compensate for this undesirable situation, it is proposed that the affected node should be allowed to transmit a new request when a conflict in the request for minislots is detected. We refer to this new approach as request-resend, and Figure 5-2 shows an example of how this procedure is carried out with the RN scheme.

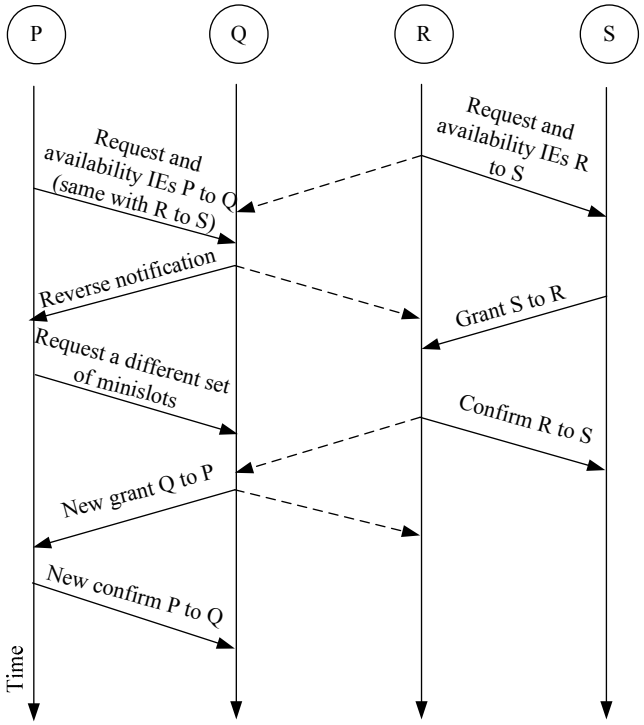



Figure 5-2 The request-resend scheme allows node P to send a new request after receiving a RN message from node Q.

Let consider the case that node P, being unaware of the minislots requested by node R, makes a request for the same set of minislots to node Q. After it receives the RN message from node Q, node P will drop its initial conflicting request and send out a new request specifying a different set of minislots. In this way, node Q will be able

to grant node P its requested minislots. Consequently, node P is able to transmit its traffic straight after confirming the grant to node Q.

Next, consider the situation where there are two additional nodes, M and N, located before node P, as shown in Figure 5-3. Let assume that this time node Q sends out a new grant straight after receiving the first request from node P, without any knowledge of the request from node M to node N. If the new set of minislots granted to node P by node Q happens to be the same as the one requested to node N by node M, this will lead to another conflict. This time, node P, being adjacent to node N, overhears the grant message from node N to node M. As a result, node P cannot confirm the new grant it has received from node Q. Now, according to the proposed request-resend scheme, instead of node Q making the decision on the set of minislots to be assigned to it, node P itself will be able to determine the set of available minislots for its own packet transmission that is not in conflict with node M. In this case, the information that node P requires to select the set of non-conflicting minislots can be derived from the grant message of node N to node M, and the RN from node Q. As such, the use of request-resend in conjunction with RN will effectively overcome the hidden node problem.

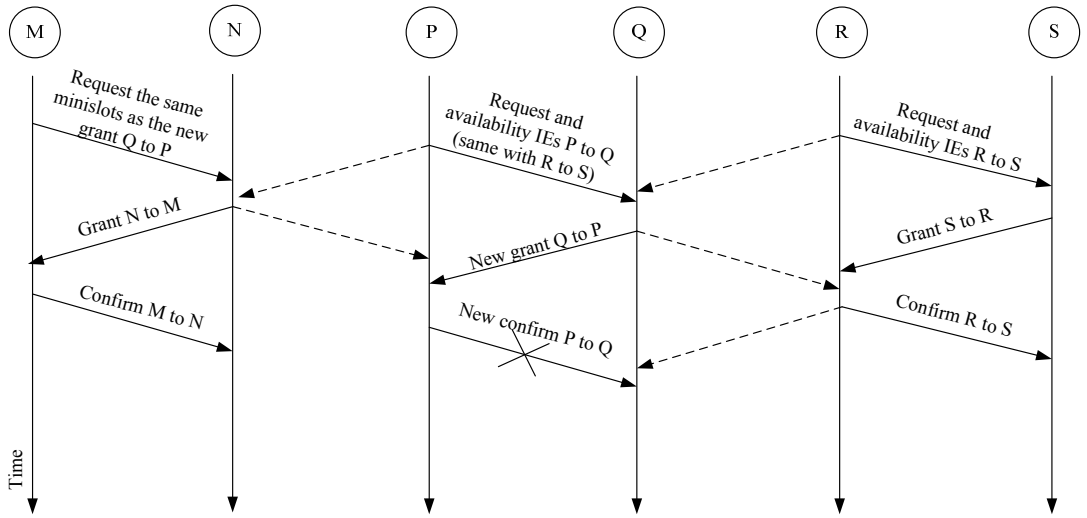



Figure 5-3 Possible hidden node problem when request-resend is not used.

### 5.3.2 Dynamic Minislot Allocation

In the event of a node failure, traffic will be rerouted bypassing the faulty node to reach the final destination via alternative paths. Hence, neighbours of the failed node are likely to have to handle a larger amount of traffic. These nodes will become bottlenecks for traffic queues to build up unless they are allowed to request more minislots to convey their data packets. Conversely, those nodes that are not involved in rerouting traffic may need to decrease their share of minislots. As such, a dynamic minislot allocation scheme is proposed for adjusting the minislot allocations according to the traffic loads serviced by individual nodes in the event of a node failure. The operation of this dynamic minislot allocation scheme involves three phases and they are described as follows.

*1. Failure detection phase*

- Under normal operation, a node will regularly broadcast MSH-DSCH messages containing its own control message transmission schedule and those of its neighbouring nodes, in addition to its request for minislots. After sending a MSH-DSCH message, the node, with the holdoff exponent value of zero, will holdoff for a period of time, which is given by

$$\begin{aligned}\text{Holdoff time} &= 2^{\text{Holdoff base + holdoff exponent}} \\ &= 2^{4+0} \\ &= 16 \text{ transmission opportunities}\end{aligned}$$

  Hence, the node will start competing for transmission opportunities after holding off for 16 transmission opportunities. As the maximum number of two-hop neighbours in a multi-hop ladder network is eight, it is expected that the node will be able to win a transmission opportunity within the next 16 transmission opportunities. As such, a node will be recognized as a failure node if it does not transmit a control message within two holdoff periods.

*2. Failure recovery phase*

- Upon detecting a node failure, the node preceding the failed node will divert its traffic via an alternative path to arrive at another node. At the mean time, it will decrease the number of minislots to be requested. The number of minislots that it should use is determined according to the procedure described in Appendix

- When a grant IE with a failure flag is received by the destined node, it will grant its upstream neighbours the same number of minislots that it has received from its immediate downstream node. Once again, the failure flag will also be included in the grant IE for passing on to the next upstream node. For the nodes who overhear the grant IE with a failure flag, they will also use the same request size and include a flag in their grant IE. In this way, the failure flag will be able to propagate to all the upstream nodes within the network that are still operating normally. Ultimately, each individual upstream node will have the required information for it to make the necessary adjustment to the number of minislots that it could grant.
- When a node receives from its upstream nodes, which under normal condition do not relay traffic to it, or they request for the number of minislots which seems larger than normal, it realises that it will have to handle additional rerouted traffic. As such, it will ignore the failure flag contained in any overheard grant IEs. Also, if it happens to have only one immediate downstream node, then it will have to increase its request size for minislots sent to this downstream node in order to accommodate the extra traffic that it has to serve. The request size may be equal to the total number of minislots requested by all its upstream nodes. On the other hand, if there are more than one immediate downstream nodes, then it is possible for the node to distribute its traffic equally among its downstream nodes. In this case, the request size made to each downstream node is equal to the total number of minislots requested by all its upstream nodes divided by the number of downstream nodes. Note that in a ladder network, each node is connected to three neighbouring nodes, so that a given node has a maximum of two immediate downstream nodes. Following this, all the subsequent downstream nodes will make use of the same request size.
- As specified in the IEEE 802.16 standard, only minislots in a continuous range can be handled by the requesting and granting nodes during a TW handshake.

*3. Return to normal operation phase*

- Upon receiving the first control message from the failed node after it has been restored to operation, its immediate neighbouring nodes will resume the use of allocated minislots meant for normal network operation. At the same time, they will reset the failure notification flag to inform other nodes in the network about the restoration of the failed node to normal operation.

## 5.4 PERFORMANCE EVALUATION OF REQUEST-RESEND AND DYNAMIC MINISLOT ALLOCATION

The proposed request-resend and dynamic minislot allocation schemes are incorporated into the IEEE 802.16 coordinated distributed scheduling and RN for use in a four-hop ladder network serving as a wireless backhaul. The performance of the backhaul network is first evaluated using the NCTUns network simulator under the condition of a single node failure. The simulation settings are as described in Section 4.4.1. Also, as a failure in any one of the two nodes on a given cross link along the two branches in the four-hop ladder network gives rise to the same number of hidden nodes, it is only necessary to examine the performance when a single node failure occurs at one of the three possible locations along either of the two parallel branches of the network, as shown in Figure 5-4.

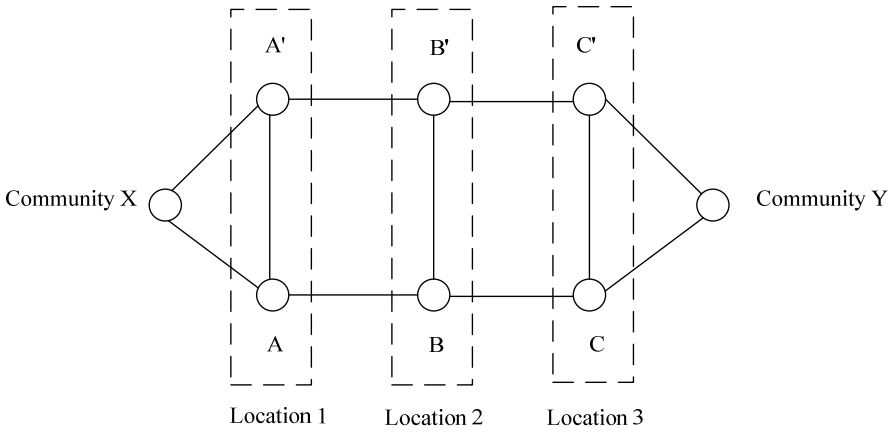



Figure 5-4 Node failure locations in the four-hop ladder topology.

The performance of the proposed request-resend scheme is evaluated using the same performance metrics adopted in Section 4.4.2, i.e., maximum achievable throughput and average end-to-end packet transmission delay. Table 5-9 shows the computer simulated throughputs and average delays obtained with the node failure occurring at the three possible node locations in the four-hop backhaul. Also presented is the theoretical throughput computed based on equation (4.4). From Table 5-9, it is observed that the adoption of the proposed request-resend and dynamic minislot allocation scheme has significantly enhanced the throughput of the four-hop backhaul when compared to the use of only the coordinated distributed scheduling and RN. More importantly, it is shown that the throughput remains constant regardless of where the node failure occurs. Also, the simulated throughput is now approaching the theoretical maximum value. These observations clearly suggest that the hidden node problem has largely been mitigated by the request-resend scheme, and better bandwidth utilisation is achieved through the use of dynamic minislot allocation. Now, with the backhaul being able to support a larger amount of traffic while maintaining no or near zero packet loss, it is expected that each node will have more packets in its buffer queue. This in turn is likely to result in an increase of average end-to-end transmission delay as shown in Table 5-9.

Table 5-9 Comparison between the throughputs obtained with and without request-resend and dynamic minislot allocation incoporated into the standard IEEE 802.16 coordinated distributed scheduling and RN.

| Failure location | Throughput (Mbps) | | | Average end-to-end delay (ms) |
|---|---|---|---|---|
| | Coordinated distributed scheduling and RN | Request-resend and dynamic minislot allocation | Theoretical throughput | |
| 1 | 8.18 | 14.28 | 15.33 | 137.95 |
| 2 | 9.11 | 14.28 | 15.33 | 137.96 |
| 3 | 6.81 | 14.28 | 15.33 | 137.93 |

Next, the above performance evaluation is extended to include ladder backhaul networks with hop counts of 2, 3, 5, and 6. The hop count referred to here is the number of hops along one of the parallel branches. The resulting computer simulated results, in terms of throughput and average end-to-end delay, are tabulated in Table 5-10.

Table 5-10 The maximum achievable throughputs and the average end-to-end transmission delays achieved with request-resend and dynamic minislot allocation in ladder networks of different hop counts operating under the condition of a single node failure.

| Number of hops | Request size | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 110 | 25.80 | 70.66 |
| 3 | 31 | 14.28 | 111.02 |
| 4 | | 14.28 | 137.95 |
| 5 | | 14.28 | 145.44 |
| 6 | 27 | 12.91 | 152.62 |

For comparison purposes, the performance of the proposed request-resend and dynamic minislot allocation schemes has also been evaluated when the ladder network is operating under either normal or failure free condition, or with a single

link failure. The simulated results are tabulated in Table 5-11 for the case of normal operation, and in Table 5-12 when operating with a single link failure. Note that the request sizes used for different hop counts are determined according to the procedures described in Section 4.4.1 and Appendix A.2 for the case of normal operation, and operating with a single link failure, respectively. As the use of the proposed request-resend and dynamic minislot allocation schemes allows similar performance to be achieved regardless of where a node or link failure occurs in the network, the throughputs and average delays as presented in Table 5-10 and Table 5-12 correspond to the ensemble average of values obtained for all possible failure locations.

Table 5-11 The maximum achievable throughputs and the average end-to-end transmission delays achieved with request-resend and dynamic minislot allocation in ladder networks of different hop counts operating under normal condition.

| Number of hops | Request size | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 55 | 25.80 | 68.79 |
| 3 | | 25.80 | 98.75 |
| 4 | 44 | 20.00 | 110.61 |
| 5 | | 20.00 | 129.15 |
| 6 | 36 | 16.02 | 140.18 |

Table 5-12 The maximum achievable throughputs and the average end-to-end transmission delays achieved with request-resend and dynamic minislot allocation in ladder networks of different hop counts operating under the condition of a single link failure.

| Number of hops | Request size | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 44 | 20.00 | 77.21 |
| 3 | 36 | 16.02 | 108.89 |
| 4 | | 16.02 | 124.87 |
| 5 | | 16.02 | 135.11 |
| 6 | 31 | 14.28 | 149.55 |

Referring to Table 5-10 to Table 5-12, as expected, the request size that could be used by a link in the wireless backhaul decreases when the hop count is increased from two to six. Consequently, as the hop count of the ladder network increases, the maximum achievable throughput becomes smaller. This is true for whether the network is operating normally, or in the presence of either a link or node failure. Moreover, a single node failure can lead to multiple link failures. As a result, many extra minislots will be needed for rerouting traffic, so that those links which are not involved in rerouting traffic will have to make use of a lesser number of minislots. This suggests why, with the exception of a 2-hop ladder network, a lowest throughput is achieved in the case of a node failure.

For a two-hop ladder network, as shown in Figure 5-5, it is possible to make use of a request size of 110 minislots in the case of normal operation and in the presence of a single node failure. On the other hand, a smaller request size of 88 minislots can only be allocated when the two-hop network is suffering from a single link failure. The use of this smaller request size also suggests a reduction in the network throughput and an increase in the average end-to-end delay when there is a single link failure occurring in the two-hop ladder network.

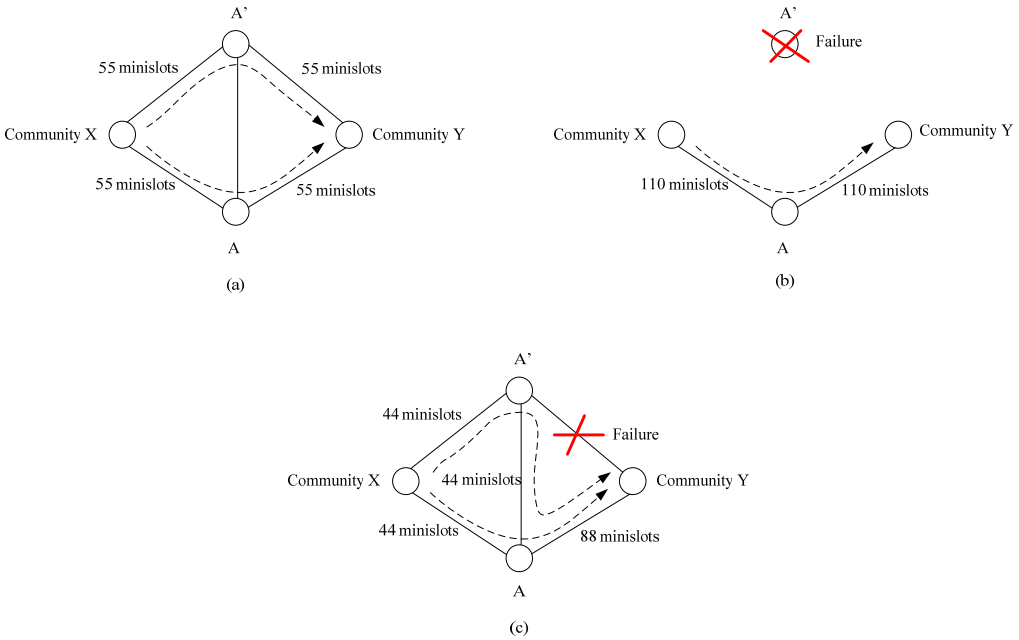



Figure 5-5 The number of minislots that could be allocated to each link in a two-hop ladder network operating under the conditions (a) failure-free; (b) a single node failure; and (c) a single link failure.

Figure 5-6 shows the throughputs obtained through the use of: (i) only the IEEE 802.16 TW handshake, (ii) TW handshake plus RN, and (iii) the proposed combination of TW handshake, RN, request-resend, and dynamic minislot allocation, in ladder networks of different hop counts. Also, included in Figure 5-6 are the theoretical maximum traffic loads that could be supported by ladder networks of different hop counts. It is interesting to note the effectiveness of the proposed request-resend and dynamic minislot allocation schemes that results in achieving throughputs which are approaching the theoretical maximum allowable traffic loads for multi-hop ladder networks.

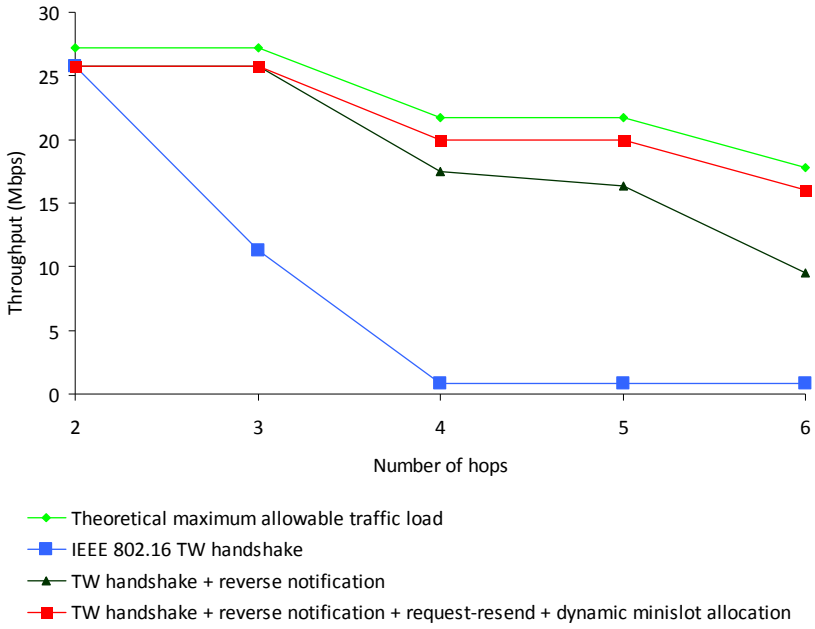



Figure 5-6 Throughputs obtained through the use of: (i) only the IEEE 802.16 TW handshake, (ii) TW handshake plus RN, and (iii) the proposed combination of TW handshake, RN, request-resend, and dynamic minislot allocation, in ladder networks of different hop counts.

So far, computer simulations have been performed based on a packet size of 1000 bytes. An attempt has been made to further increase the maximum achievable throughput through the use of a larger packet size. However, simulated results, presented in Figure 5-7, show that the request size used tends to restrict the amount of traffic that can be transmitted at a given time.

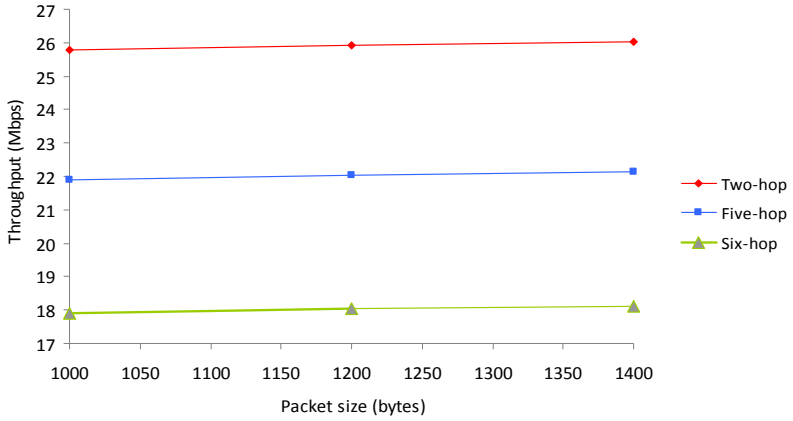



Figure 5-7 Throughputs of the two-hop, five-hop, and six-hop ladder backhauls obtained for three different packet sizes. Note that the maximum transmission unit (MTU) of Ethernet is 1500 bytes.

As a result, any throughput gain based on the use of packet size above 1000 bytes is marginal. Similar results are also observed for the three-hop and four-hop ladder networks. These results have been obtained by adopting the proposed request-resend and dynamic minislot allocation schemes.

Now, with the hidden node problem associated with a multi-hop ladder network largely overcomes through the use of request-resend and dynamic minislot allocation, the next step is to examine the influence of the buffer size adopted at each node on the maximum amount of traffic load that could be handled by the network. In this case, it is assumed that packet loss caused by buffer overflow remains the main factor governing the allowable maximum traffic load. On the other hand, a quite different scenario is associated with the case that makes use of the original IEEE 802.16 coordinated distributed scheduling and RN. For the latter, the severe hidden node problem encountered by the original IEEE 802.16 coordinated distributed scheduling protocol is the main cause hindering the ability of a node to transmit its incoming traffic in a timely fashion. As a result, the traffic queue soon builds up and causes its buffer to overflow. Although the use of RN has partially improved the situation, there remain special circumstances in which data transmission is still being restricted at some nodes due to the hidden node problem. The above observations suggest that the use of a longer buffer is likely to give rise to a longer end-to-end transmission delay.

For comparison, Table 5-13 presents the average end-to-end delays obtained through the use of (i) the original IEEE 802.16 coordinated distributed scheduling, (ii) RN, and (iii) request-resend and dynamic minislot allocation in ladder networks of different hop counts. It is to be noted that the slightly longer delays associated with case (iii) have been achieved for higher throughputs as compared to case (ii). Nonetheless, the average transmission delays obtained for ladder networks up to a hop count of 6 remain below the 150 ms limit generally accepted for services involving delay sensitive real-time traffic.

Table 5-13 Average end-to-end transmission delays obtained through the use of (i) IEEE coordinated distributed scheduling, (ii) RN, and (iii) request-resend and dynamic minislot allocation in wireless ladder backhauls of hop counts up to six. The buffer size used is 1000 bytes.

| Number of hops | Request size (minislots) | Average end-to-end packet transmission delay (ms) | | |
|---|---|---|---|---|
| | | IEEE 802.16 coordinated distributed scheduling | RN | Request-resend and dynamic minislot allocation |
| 2 | 55 | 70.23 | 69.67 | 68.79 |
| 3 | | 24.16 | 94.15 | 98.75 |
| 4 | 44 | 317.73 | 104.78 | 110.61 |
| 5 | | 365.75 | 83.45 | 129.15 |
| 6 | 36 | 384.04 | 53.71 | 140.18 |

The influence of buffer size on the average end-to-end transmission delay for a ladder network that incorporates the request-resend and dynamic minislot allocation schemes is illustrated in Figure 5-8. As expected, the average delay tends to decrease when a smaller buffer size is used. The delay initially decreases gradually for buffer size down to 200 packets, and is followed by a rapid fall in delay thereafter. This is true for ladder networks of different hop counts. Although an 18 % reduction in average delay could be achieved by adopting a buffer size of 200 packets instead of 1000 packets for a 5-hop ladder network, this is achieved with a corresponding decrease in throughput of 19 %, as shown in Figure 5-9. Moreover, the throughput of a ladder network is not significantly affected by the use of buffer sizes ranging from

200 to 1000 packets. Observations from Figure 5-8 and Figure 5-9 suggest that multi-hop ladder networks, that incorporate the use of request-resend and dynamic minislot allocation, are able to operate well over a large range of buffer sizes.

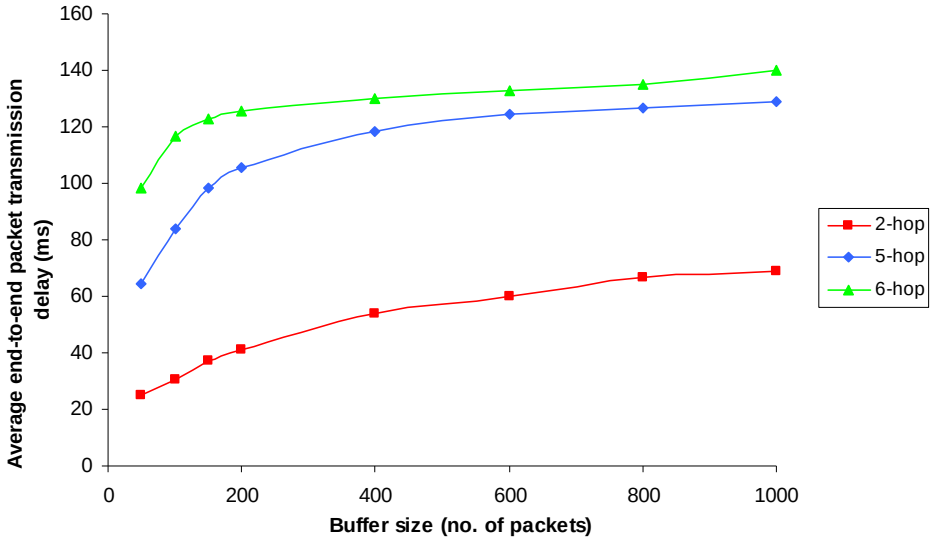



Figure 5-8 Variations of average end-to-end packet transmission delay with buffer size used in wireless ladder backhauls of three different hop counts. The use of request-resend and dynamic minislot allocation is assumed.

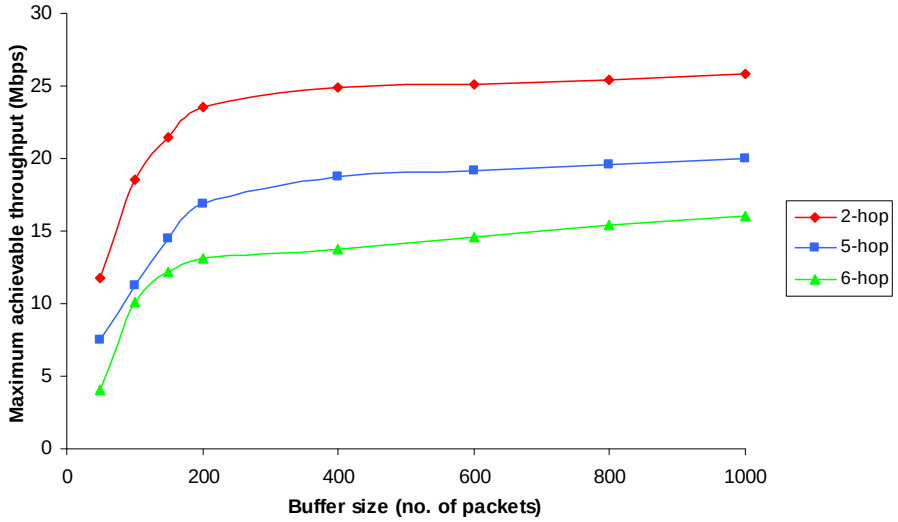



Figure 5-9 Maximum achievable throughputs obtained as a function of the buffer size used for ladder networks with hop counts of two, five and six. The use of request-resend and dynamic minislot allocation is assumed.

Thus far, all the results have been obtained by applying the appropriate traffic load to meet the no or near zero packet loss criterion for determining the maximum traffic load that could be supported by the network. In addition, Figure 5-10 and Figure 5-11 show how the packet drop rate and the average end-to-end packet transmission delay vary with traffic loads for the two-hop, five-hop, and six-hop ladder networks.

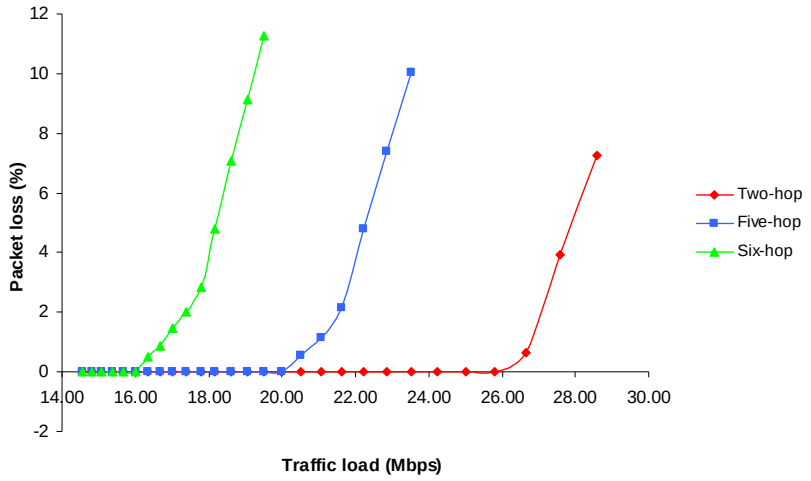



Figure 5-10 Percentage packet loss as a function of traffic load for the two-hop, five-hop, and six-hop ladder networks.

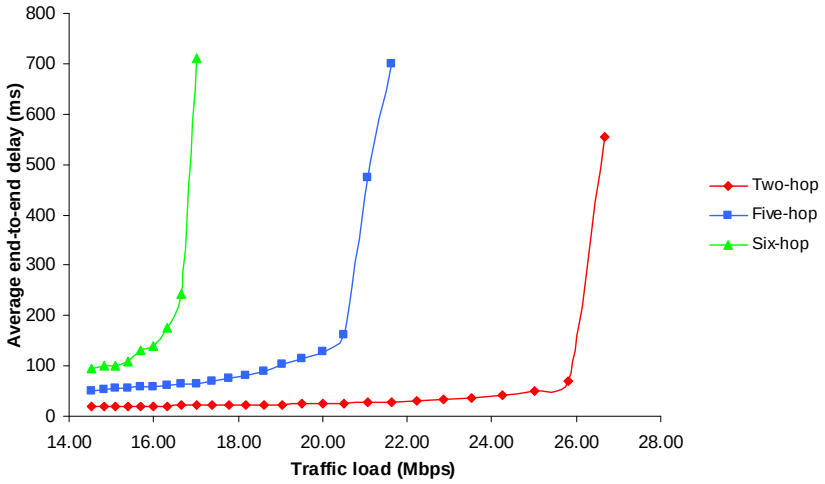



Figure 5-11 Average end-to-end packet transmission delay as a function of traffic load for the two-hop, three-hop, and six-hop ladder networks.

From Figure 5-10 and Figure 5-11, it is observed that all the three ladder networks show an initial gradual increase in both the packet drop rate and average transmission

delay with increasing traffic loads. However, as expected, when the traffic load approaches a certain threshold, referred here as the maximum allowable traffic load, beyond which both the packet drop rate and average transmission delay increase rapidly. In fact, as shown in Figure 5-12, there is a deterioration in throughput when a given network is loaded with traffic beyond the maximum allowable value. Although not shown here, similar observations have also been made for the three-hop and four-hop ladder networks.

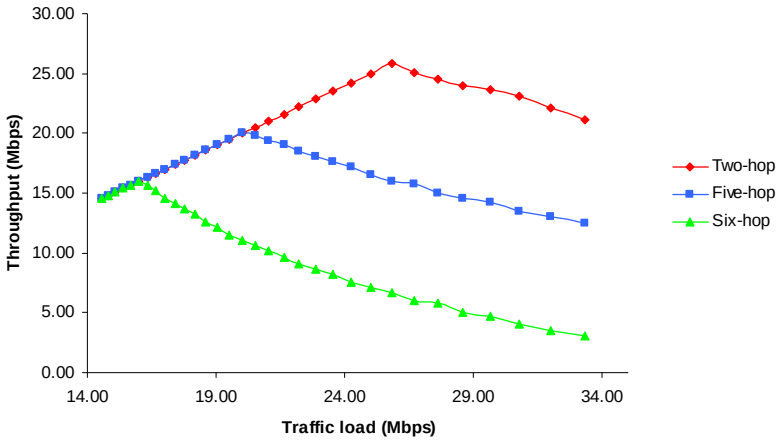



Figure 5-12 Throughput of the two-hop, five-hop, and six-hop ladder network when the traffic load is increased beyond the maximum allowable value.

## 5.5 PERFORMANCE ACHIEVED WITH A NETWORK CONSISTING OF TWO PARALLEL PATHS WITH THE SAME HOP COUNT

As discussed in Section 4.2.2, an obvious way to improve the failure resilience of a chain network is to duplicate the chain of nodes to result in a two parallel path network, as shown in Figure 4-3(a). The operation of such a network is still able to sustain a single node or link failure occurring in either one of its two branches. Moreover, because of the absence of any cross links, the number of neighbouring nodes within two hops of a given node becomes smaller in the case of this two parallel path network. As such, during normal operation, each of the two parallel paths is able to make use of a larger request size. This is likely to lead to a higher throughput than that of a ladder network of the same hop count. On the other hand, if one of the two branches suffers from either a node or link failure, then only half of the network remains operational, with a consequent significant loss of throughput.

Therefore, it remains interesting to examine the performance of such a two path network operating with the use of the IEEE 802.16 coordinated distributed scheduling in conjunction with RN, request-resend, and dynamic minislot allocation. Computer simulated results obtained using the same simulation settings as described in Section 4.4.1 are tabulated in Table 5-14 and Table 5-15 for the network operating normally and with single failure, respectively.

Table 5-14 The maximum achievable throughputs and average end-to-end packet transmission delays obtained for the two parallel path networks of five different hop counts operating under normal condition.

| Number of hops | Request size | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 110 | 49.97 | 53.59 |
| 3 | 55 | 25.80 | 100.17 |
| 4 | | 25.80 | 107.55 |
| 5 | | 25.80 | 124.55 |
| 6 | | 25.80 | 136.77 |

Table 5-15 The maximum achievable throughputs and average end-to-end packet transmission delays obtained for the two parallel path networks of five different hop counts operating in the presence of a node or link failure.

| Number of hops | Request size | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 110 | 25.80 | 70.66 |
| 3 | 73 | 17.39 | 98.55 |
| 4 | | 17.39 | 105.90 |
| 5 | 55 | 13.11 | 129.91 |
| 6 | | 13.11 | 138.66 |

When compared with the results obtained for the ladder networks of comparable hop count, as shown in Tables 5-10 to 5-12, the throughputs achieved by the two parallel path network under normal condition are indeed higher as expected. When there is a single node or link failure, both networks achieve similar throughputs. Moreover,

when a single node or link failure occurs in both branches, the two parallel path network will cease operation while the ladder network will likely remain operational, albeit with a reduction in throughput.

## 5.6 PERFORMANCE EVALUATION OF REQUEST-RESEND AND DYNAMIC MINISLOT ALLOCATION WITH BIDIRECTIONAL TRAFFICS

In practice, traffics in a wireless backhaul are flowing in both directions. This means that it is crucial to also examine how a ladder network performs in the presence of bidirectional traffics. For the performance evaluation, the computer simulation settings, as described in Section 4.4.1, are again adopted. In this case, traffic is being relayed in the direction from Community X to Community Y, as well as from Y to X. Also, the request size used is determined according to the procedure presented in Appendix B. This request size is meant for those links which are not involved in rerouting traffic during a node or link failure. The simulated maximum achievable throughputs and average transmission delays are presented in Table 5-16, Table 5-17 and Table 5-18 for the cases when the network is operating normally, with a single node failure, and with a single link failure, respectively.

Table 5-16 Maximum achievable throughputs and average transmission delays obtained for the ladder networks of different hop counts operating normally with bidirectional traffics. The use of request-resend and dynamic minislot allocation is assumed.

| No. of hops | Request size | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 27 | 12.70 | 12.70 | 85.12 | 87.34 |
| 3 | | 11.38 | 11.38 | 120.41 | 121.58 |
| 4 | 22 | 10.39 | 10.39 | 158.78 | 158.95 |
| 5 | | 10.26 | 10.26 | 160.51 | 162.01 |
| 6 | 20 | 7.96 | 7.96 | 202.02 | 196.76 |

Table 5-17 Maximum achievable throughputs and average transmission delays obtained for the ladder networks of different hop counts operating with bidirectional traffics in the presence of a single node failure. The use of the request-resend and dynamic minislot allocation is assumed.

| No of hops | Request size | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 55 | 13.11 | 13.11 | 72.69 | 73.43 |
| 3 | 20 | 9.04 | 9.04 | 132.88 | 132.45 |
| 4 | 18 | 7.94 | 7.94 | 167.71 | 169.51 |
| 5 | | 6.97 | 6.97 | 172.91 | 174.77 |
| 6 | 17 | 6.54 | 6.54 | 214.74 | 216.69 |

Table 5-18 Maximum achievable throughputs and average transmission delays obtained for the ladder networks of different hop counts operating with bidirectional traffics in the presence of a link failure. The use of the request-resend and dynamic minislot allocation is assumed.

| No of hops | Request size | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 22 | 10.39 | 10.39 | 95.39 | 95.66 |
| 3 | 20 | 9.04 | 9.04 | 131.79 | 131.54 |
| 4 | | 8.67 | 8.67 | 161.11 | 162.23 |
| 5 | | 7.54 | 7.54 | 170.55 | 171.91 |
| 6 | 18 | 6.67 | 6.67 | 213.54 | 214.45 |

In order to support bidirectional traffic flows, the total number of available minislots will now have to be divided into two equal sets of minislots, one for each direction. As such, the request size for an individual link serving one direction is half or slightly smaller than that used to support a unidirectional traffic flow. This in turn suggests that the achievable throughput in each direction of the wireless backhaul will be roughly half of that for the case with unidirectional traffic. This remains true irrespective whether the network is operating under a normal or faulty condition.

Also, it is noted that the introduction of bidirectional traffic transmission will increase the number of scenarios whereby the hidden node problem could occur in a ladder network. One such situation is shown in Figure 5-13, which involves node P and node Q that are within two-hop of one another but transmit in different directions making requests for the same minislots to node Q. In this case, node P would have to send a new request with the use of the request-resend scheme. The occurrences of the hidden node problem associated with bidirectional traffics would decrease the throughput in ladder networks of different hop counts. Moreover, the scenario of Figure 5-13 will not be a problem if node Q is able to send a grant IE back to node P before node R makes its request. Under this situation, node R will be informed of the minislots granted to node P and it will then avoid making a request for the same minislots. A more positive way to avoid the occurrence of such a scenario is to make use of a different frequency channel for node R to transmit its data. This approach will be further investigated in Chapter 6.

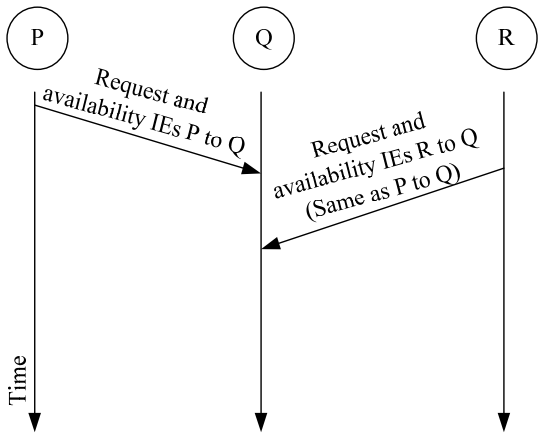



Figure 5-13 A possible scenario that hidden node problem could occur in the case of bidirectional traffic transmission.

From Tables 5-16 and 5-17, it is a bit surprising to observe that the throughput for a two-hop ladder network with a single node failure is indeed larger than that under normal network operation. For this particular case, when an intermediate node fails, then the ladder network is reduced to a single 2-hop chain network, as shown in Figure 5-14(a). Under this situation, each node will be able to obtain 55 minislots to transmit data in each direction. On the other hand, when the network is operating

normally, the number of minislots that could be allocated to each node for data transmission is reduced to only 27, as shown in Figure 5-14(b). This explains why a lower throughput is obtained when the 2-hop ladder network is operating normally. For completeness, Figure 5-14(c) shows the minislots allocation when a link failure occurs in a 2-hop ladder network.

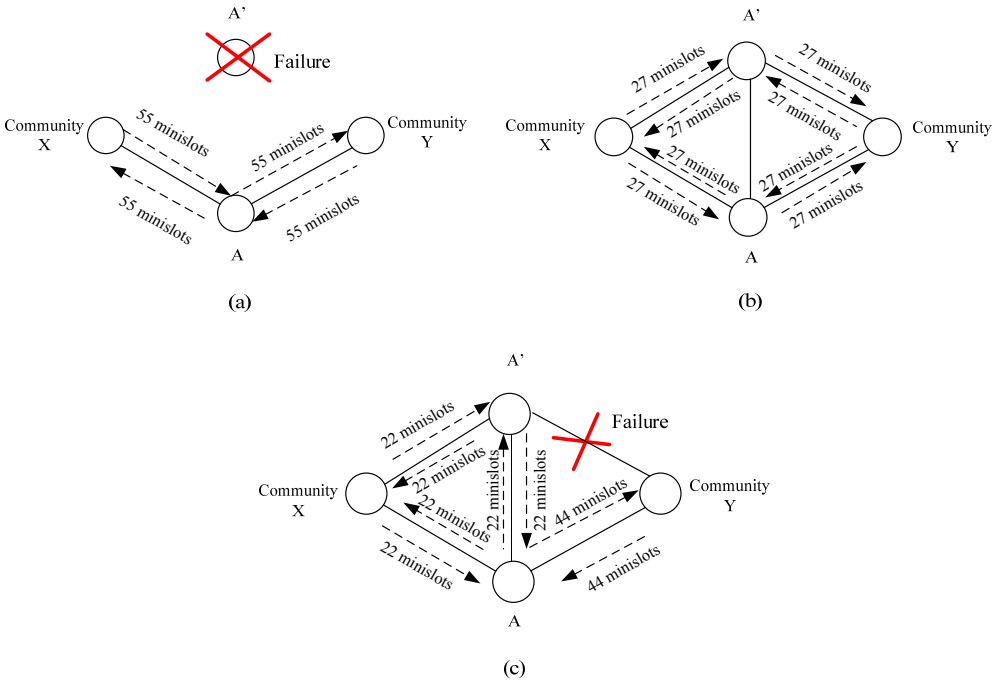



Figure 5-14 Allocations of minislots for bidirectional traffic transmissions in a 2-hop ladder network operating under: (a) a single node failure; (b) normal condition; and (c) a single link failure.

Next, the same performance evaluation has also been carried out for the two parallel path network. The resulting simulated maximum achievable throughputs and average end-to-end packet transmission delays obtained for bidirectional traffic transmissions when the network is operating normally and in the presence of a single node or link failure are tabulated in Table 5-19 and Table 5-20, respectively.

Table 5-19 Maximum achievable throughputs and average end-to-end packet transmission delays obtained for the two parallel path networks operating under bidirectional traffics during normal operating condition.

| No. of hops | Request size | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 55 | 25.80 | 25.80 | 69.54 | 67.82 |
| 3 | 31 | 14.28 | 14.28 | 112.57 | 111.22 |
| 4 | | 14.02 | 14.02 | 134.55 | 137.11 |
| 5 | | 13.85 | 13.85 | 151.01 | 153.77 |
| 6 | | 13.58 | 13.58 | 167.50 | 171.54 |

Table 5-20 Maximum achievable throughputs and average end-to-end packet transmission delays obtained for the two parallel path networks operating under bidirectional traffics in the presence of a single node or link failure.

| No. of hops | Request size | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 55 | 13.11 | 13.11 | 72.98 | 73.05 |
| 3 | | 13.05 | 13.05 | 115.82 | 119.44 |
| 4 | 44 | 10.00 | 10.00 | 138.33 | 139.69 |
| 5 | | 9.52 | 9.52 | 146.77 | 147.25 |
| 6 | 31 | 7.37 | 7.37 | 172.22 | 168.92 |

When comparing the above results with those obtained for the ladder networks as presented in Tables 5-16 to 5-18, it is observed that the resulting throughputs follow the same trend as that for the case of unidirectional traffic. For example, a higher throughput is obtained with the two parallel path network compared to the ladder network. This difference is more significant in the case of normal operating condition. Under the normal condition, the two parallel path network may be considered as two independent chain networks operating at the same frequency. This suggests that the two chain networks must have their nodes separated sufficiently far apart so that they do not affect each other's minislot allocation. As such, the two parallel path network also encounters a less severe hidden node problem compared

with the ladder network. On the other hand, the former will not survive a single node or link failure occurring in both branches of the network.

In general, the decrease in the request size of individual nodes servicing bidirectional traffic has reduced the amount of traffic that can be transmitted in a given period of time. Consequently, packets will wait in the buffer queue of each intermediate node for a longer period of time before they are transmitted. This results in an increase of average end-to-end packet transmission delay for both the two parallel path network and the ladder network when the networks are engaged in bidirectional traffic transmissions.

## 5.7 SUMMARY

The RN scheme, discussed in the last chapter, for mitigating the hidden node problem in a multi-hop ladder network is shown to be less effective when the network suffers a node failure. This is caused by the tendency of a RN message to arrive at the destined node only after it has already made a request. Also, the presence of a node failure will result in changes in the number of hidden nodes associated with the remaining network nodes. This then results in throughput which becomes dependent on the actual location of the failing node. Two new schemes, referred to as the request-resend and dynamic minislot allocation schemes, when operating in conjunction with RN, have been shown not only to overcome the above shortcoming, but also to enhance the network throughputs. It is observed that the throughputs achieved during normal condition are now very close to the theoretical maximum traffic loads that could be supported by ladder networks of different hop counts. Furthermore, computer simulations show that the resultant throughputs achieved through the use of packet size of 1000, 1200 and 1400 bytes are very similar. In addition, the use of a smaller buffer size in each node tends to lower the average end-to-end delay. The change is gradual for buffer sizes in the range of 200 to 1000 packets. Below a buffer size of 200 packets, smaller delays are observed but these have been achieved with corresponding decrease in throughput.

A subsequent performance comparison between the ladder network and the two parallel path network shows that the latter is able to achieve a larger throughput during normal operating condition. This is due to the fact that a two parallel path network may be considered as two independent chain networks, so that each node along the separate chain will have a smaller number of one-hop and two-hop neighbouring nodes, compared with a ladder network. As such, a given node in the two parallel path network is able to make use of a larger request size to deliver a larger throughput. However, when a node or link fails in one of the two parallel paths, the throughput will be halved. In the case of a node or link failure occurring in both branches, the two parallel path network will cease to operate while the ladder network is likely to survive such failures by rerouting traffic around the failed nodes using the cross links between the two branches.

In order to support traffic transmissions in both directions, each link in the network has to request two different sets of minislots, one for each direction. Since there is a fixed number of minislots available for a single channel operation, this means that each bidirectional link will have to share a smaller request size, thus leading to a lower throughput compared with unidirectional transmission in both the ladder and two parallel path networks. Also, the average end-to-end packet transmission delay is increased as the amount of traffic that could be transmitted at a given time is reduced. Moreover, bidirectional traffic transmissions also give rise to additional scenarios relating to the hidden node problem. It will be shown in the next chapter that such hidden node scenarios can be resolved through the use of dual frequency operation.

# CHAPTER 6

# TWO-CHANNEL TWO-TRANSCEIVER IEEE 802.16 WIRELESS BACKHAUL

## 6.1 INTRODUCTION

In Section 5.4, it is shown that the hidden node problem in a multi-hop WiMax wireless network has to a large extent been overcome by incorporating the proposed request-resend scheme into the initial channel resource request procedure. This in combination with the use of dynamic minislot allocation has significantly improved the minislot utilisation efficiency. As a result, the performance of the wireless backhaul, based on the proposed ladder network topology, is significantly enhanced when operating under normal conditions. However, due to the fact that only a limited number of minislots are available for distributing to individual nodes, when operating on a single frequency channel, the throughput of the wireless backhaul is somewhat constrained. This is especially so for the backhaul of more than three hops. In fact, Table 5-11 shows that a maximum throughput of 25.8 Mbps is achieved for a backhaul with only two to three hops. Also, it has been identified in Section 5.6 that the hidden node problem is even more acute when the backhaul is carrying bidirectional traffics. It is envisaged that the use of an additional frequency channel may assist in further alleviating the hidden node problem, as well as making more minislots available to enhance the throughput of the backhaul.

In Section 6.2, a two-channel two-transceiver IEEE 802.16 distributed channel assignment (TTDCA) scheme is described. Its performance, in terms of throughput and average delay, is then evaluated and presented in Section 6.3. Also, the influence on the network performance due to the use of different holdoff exponent values is examined. A modification is then made to the TTDCA scheme to enable the second transceiver to operate during the control subframes, as described in Section 6.4. The computer simulated results obtained with this modified TTDCA scheme are presented in Section 6.5, for ladder networks operating under normal condition, and

in the presence of a link or node failure. In the case of bidirectional traffics, the results are presented in Section 6.6.

## 6.2 TWO-CHANNEL TWO-TRANSCEIVER DISTRIBUTED CHANNEL ASSIGNMENT

When a ladder backhaul is operating using a single frequency channel, the concurrent transmissions of nodes located within the interference range will give rise to transmission collisions. To avoid this situation, different frequency channels may be assigned to these nodes to allow them to transmit simultaneously without causing collisions. At the same time, more minislots will be made available for data transmission. Moreover, to enable concurrent transmissions and receptions by individual nodes, they will need to be equipped with multiple transceivers, which lead to greater expense. This chapter examines how the use of two transceivers per node could enhance the performance of a multi-hop wireless ladder backhaul, in terms of throughput and average latency.

Frequency channel allocation in a backhaul network can be carried out either centrally or in a distributed fashion. Most of the published algorithms for channel and minislot assignments associated with IEEE 802.16 networks are based on centralised scheduling, in which a base station is assigned the responsibility for allocating minislots to individual subscribing stations upon their requests [85-93]. However, such approaches tend to incur long connection set up time as the base station needs to gather resource requests from all the subscribing stations before allocating network resources to them [61]. A more flexible approach is to let each node in the network to coordinate its own transmission schedule with its two-hop neighbouring nodes without having to interact with a particular base station. This is commonly referred to as distributed channel assignment, which may require less connection set up time. This latter option becomes more attractive for the multi-hop wireless backhaul proposed in this thesis.

In [80, 81], a multichannel single-transceiver distributed channel assignment algorithm is proposed. With this algorithm, each node will tune to a common

channel, say channel 1, during the control subframe period to receive control messages transmitted by its immediate neighbouring nodes. In this way, a node will be able to learn about those minislots that have already been taken and thus avoid making a request for them. Also, the node will take note of the availability statues of individual minislots in a given frame and channel, and record them in a three-dimensional bit map, as shown in Figure 3-14. It then makes use of this bit map to search for any free minislots using the procedure as illustrated in Figure 3-15. For example, it starts by randomly select a channel, and then begins the search in a sequential manner starting from the first minislot in the first frame of the channel. If no free minislot is available in that channel, the node will continue its search for available minislots in the next channel. In the case it fails to find any available minislots in the first frame after searching in all the channels, the node will move on to continue searching in the second frame. This process continues until either the node obtains all the bandwidth it requires, or it has completed its search for available minislots in all the frames and channels.

Now, let consider the case which adopts two frequency channels, say CH1 and CH2. Accompanied with these two channels are their respective frames and corresponding minislots. Moreover, these frames are running at the same time step, i.e., they are synchronous. The introduction of an extra channel should, in principle, double the number of minislots which could be allocated for data transmission. But as each node is equipped with only a single time division duplex (TDD) radio transceiver in the schemes outlined in the last paragraph, a given node can only either transmit or receive on a particular channel at any one time. This could give rise to the situation discussed in the following example. If a given node, say node X, after searching through its bit map, discovers that a particular minislot in frame 1 of CH1 is free, it then makes a request for this particular minislot to transmit to its neighbour, say node Y. Upon receiving the request, node Y goes through its bit map record, and discovers that the same numbered minislot in frame 1 of CH2 has already been taken. This means that at the time of this minislot, node Y has already committed itself to transmit or receive on CH2. Since it is not possible for a node to operate simultaneously on both frequencies, node Y will then refuse to grant node X its request. This scenario suggests that a receiving node will only be able to grant a request for a particular minislot if both the same numbered minislots associated with

CH1 and CH2 are free. As such, even though there are more minislots available due to the use of two frequency channels, not all of them could be allocated for data transmission at a given time.

On the other hand, by equipping each node with two TDD transceivers, it is possible to overcome the above shortcoming associated with the single transceiver case. Based on the same example as discussed in the last paragraph, node Y has prior commitment to a certain minislot on CH2 at the time when it received a request from node X for granting the use of the same numbered minislot but on CH1. Now, with the use of an additional transceiver, node Y is able to transmit or receive on both frequencies at the same time. This means that node Y is able to grant node X its requested minislot, knowing that it will be possible for it to later receive the data conveyed in this minislot on CH1. Consequently, it is proposed that the multichannel single-transceiver distributed channel assignment algorithm in [80, 81] be modified by equipping a node with one additional transceiver. In particular, a two-channel two-transceiver distributed channel assignment (TTDCA) algorithm is considered in this chapter. With this algorithm, all the nodes will make use of only one transceiver to tune to the common channel, say channel 1, during the control subframe to exchange control messages between immediate neighbouring nodes. During data subframe, each node will then use both the two transceivers to transmit and receive data on CH1 and CH2. At the same time, the availability statues of all the individual minislots will be recorded in a three-dimensional bit map. The same procedure, as proposed in [80, 81], is adopted here to search the bit map for free minislots. The use of two transceivers instead of one in each node has enhanced the flexibility of utilising any available free minislots. By referring back to the example shown in Figure 3-15, it is shown that the granting node is now able to allocate minislot 4 from frame 2 in CH2 to its requesting node even though it overlaps in time with the already committed minislot 4 from the same frame in CH1.

In an attempt to lessen the impact of the hidden node problem on the ladder network, it is proposed that the links at the same hop count level along the two parallel branches be pre-assigned to operate on the same channel, as shown in Figure 6-1(a). Furthermore, the two channels, CH1 and CH2, are interchanged at alternate hop count levels. It is to be noted that a cross link will operate on the same transmit

frequency as the two nodes that it is attached to. For example, the cross links across B'B and C'C will be operating on CH1 and CH2 respectively. With this arrangement, it is possible to resolve the hidden node problem encountered by those nodes that are handling bidirectional traffics in Figure 5-13. Also, organising the two frequency channels in this way will reduce the number of nodes that appear hidden from a given node. For example, suppose we consider all the nodes are transmitting in the same direction, as shown in Figure 6-1(b). Let node C' transmits on CH2 via link $L_7$, which will normally result in node C' having the largest collision domain set (CDS) value, if all the other nodes are also transmitting on the same channel. In this case, nodes A', B, D, and E' will become the hidden nodes of node C'. Now, by arranging nodes B and D to transmit on a different channel, via links $L_6$ and $L_{10}$, respectively, they will no longer be considered as the hidden nodes of node C'. It is proposed that the TTDCA algorithm will operate in conjunction with reverse notification (RN) and request-resend to further improve the performance of the ladder network.

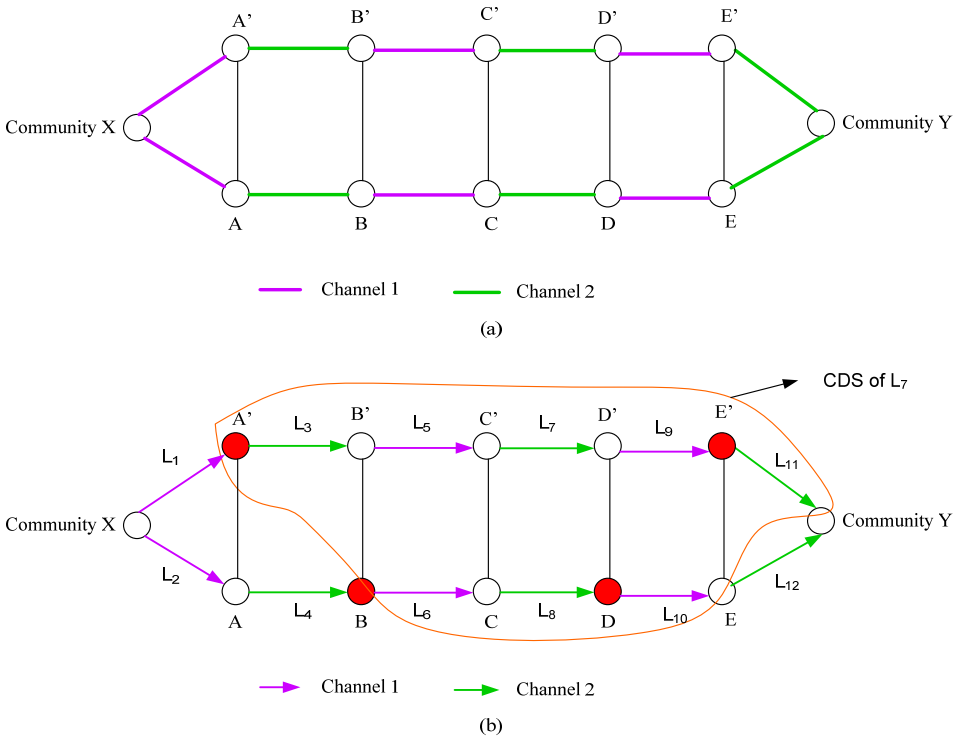



Figure 6-1 (a) Channel allocations to individual links in a six-hop ladder network. Note that the frequency channels for the cross links are not defined during normal condition as they are not used to route traffic; (b) The highlighted nodes, A', B, D, and E', appear hidden from node C', which has the largest CDS value.

## 6.3 PERFORMANCE OF THE TWO-CHANNEL TWO-TRANSCEIVER DISTRIBUTED ASSIGNMENT ALGORITHM

In this section, the performance of the TTDCA algorithm operating in conjunction with RN and request-resend in a ladder network is evaluated based on the simulation settings described in Section 4.4.1. The request size for an individual link is calculated following the procedure presented in Appendix C.1.

Table 6-1 shows the resultant maximum achievable throughputs and average end-to-end packet transmission delays achieved for ladder networks of 2 to 6 hops. For comparison purposes, the corresponding throughputs and delays obtained through the use of a single frequency channel, which has a total bandwidth equal to half of that of the two-channel case, are also included in Table 6-1. As expected, the adoption of the TTDCA algorithm has effectively alleviated problems associated with hidden nodes, as well as maximising the opportunity for nodes to transmit concurrently. As a result, a uniform throughput of 50 Mbps is achieved irrespective of the hop count of the ladder network through the use of TTDCA. This throughput is at least double those obtained with the single frequency channel case. Also, with TTDCA, individual nodes are able to handle a larger amount of traffic, and this gives rise to slightly reduced average end-to-end packet transmission delays.

Table 6-1 Maximum achievable throughputs and average end-to-end packet transmission delays for ladder networks of different hop counts operating with TTDCA in conjunction with RN and request-resend.

| Number of hops | Maximum achievable throughput (Mbps) | | Average end-to-end transmission delay (ms) | |
|---|---|---|---|---|
| | TTDCA | Single channel | TTDCA | Single channel |
| 2 | 50.00 | 25.80 | 57.49 | 68.79 |
| 3 | 50.00 | 25.80 | 92.17 | 98.75 |
| 4 | 50.00 | 20.00 | 100.24 | 110.61 |
| 5 | 50.00 | 20.00 | 116.47 | 129.15 |
| 6 | 50.00 | 16.02 | 135.97 | 140.18 |

As stated in [66, 74], the use of a smaller holdoff base value can reduce the time taken to complete a three-way handshake. It then becomes possible to make use of a smaller holdoff base value to reduce the average packet transmission delay without sacrificing the maximum achievable throughput. Moreover, it is proposed in [66] that equation (3.6) may be used to determine a suitable holdoff exponent (*hexp*) for a given node, with a holdoff base of zero, according to its number of neighbouring nodes, which are within two-hop away. For example, consider a four-hop ladder network as shown in Figure 6-2. The computed values of *hexp* for the individual nodes based on their corresponding numbers of neighbouring nodes, which are within two-hop away, are tabulated in Table 6-2.

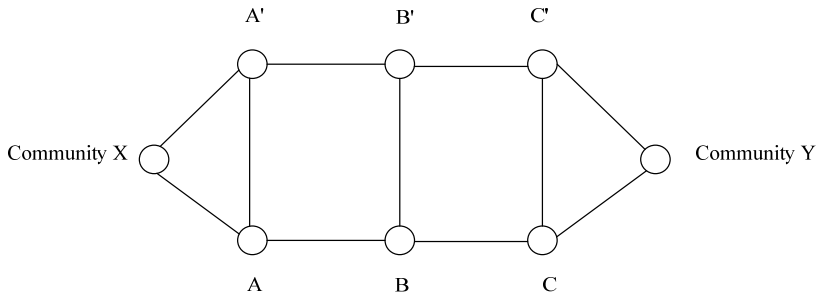



Figure 6-2 Four-hop ladder network.

Table 6-2 The computed *hexp* value of each individual node in a 4-hop ladder network together with its number of neighbouring nodes, which are within two-hop away.

| Node | Number of neighbours | Holdoff exponent, *hexp* |
|---|---|---|
| X | 5 | 2 |
| A | 6 | 2 |
| A’ | 6 | 2 |
| B | 8 | 3 |
| B’ | 8 | 3 |
| C | 6 | 2 |
| C’ | 6 | 2 |
| Y | 5 | 2 |

It is interesting to note that with the exception of node B and B', the same *hexp* value of two is found to be suitable for use in the nodes. For node B and B', although the calculated *hexp* value is 3, they can also utilise the *hexp* value of 2 as some of their neighbours can transmit simultaneously without causing collisions, such as node X and Y which are more than two hops away from each other. As such, it is expected that the adoption of the *hexp* value of 2 in a 4-hop ladder network should yield the lowest transmission delay. Although not shown here, it has been found that the same *hexp* value of two is also suitable for use in ladder networks of other hop counts, as their nodes have the similar number of neighbours.

Four integer values of *hexp*, ranging from 1 to 4, are then used to determine the maximum achievable throughputs and average end-to-end packet transmission delays of ladder networks of hop counts from 2 to 6, operating with TTDCA in conjunction with RN and request-resend. Computer simulations are carried out using the same simulation settings as given in Section 4.4.1, and the holdoff base is set to zero. The simulated results obtained are tabulated in Table 6-3.

Table 6-3 Maximum achievable throughputs and average end-to-end delays obtained for ladder networks of five different hop counts with four different *hexp* values using the TTDCA algorithm.

| No of hops | Maximum achievable throughput (Mbps) | | | | Average end-to-end delay (ms) | | | |
|---|---|---|---|---|---|---|---|---|
| | *hexp* = 4 | *hexp* = 3 | *hexp* = 2 | *hexp* = 1 | *hexp* = 4 | *hexp* = 3 | *hexp* = 2 | *hexp* = 1 |
| 2 | 50 | 50 | 50 | 50 | 57.49 | 35.34 | 25.36 | 26.60 |
| 3 | 50 | 50 | 50 | 50 | 92.17 | 68.54 | 40.57 | 41.50 |
| 4 | 50 | 50 | 50 | 50 | 100.24 | 78.23 | 55.27 | 57.35 |
| 5 | 50 | 50 | 50 | 50 | 116.47 | 92.61 | 74.96 | 76.15 |
| 6 | 50 | 50 | 50 | 50 | 135.97 | 123.26 | 89.95 | 91.55 |

Indeed, the transmission delay of the ladder network initially reduces with a decrease in the value of *hexp* used. A minimum delay is obtained when the value of *hexp* is equal to two. The use of a smaller *hexp* value than two will result in a larger delay. This is true for all the five ladder networks considered. The above observation verifies that the *hexp* value calculated using equation (3.6) is in fact optimal. It is also

noted that the choice of the *hexp* value does not affect the maximum achievable throughput.

## 6.4 USE OF SECOND TRANSCEIVER DURING CONTROL SUBFRAME

According to the TTDCA algorithm, only one transceiver in each node is used to tune to the control subframe on the common channel. Now, instead of letting the second transceiver be left idle, it is therefore interesting to investigate whether the use of this transceiver during the period of the control subframe could further improve the performance of the ladder network.

There are two ways of making use of the second transceiver during the control subframe period on the second channel. One way is to modify the control subframe of this channel to support data transmission. However, in view of the fact that the control subframe only occupies 7% of a frame, any gain in throughput is not likely to be significant. For this reason, this will not be further considered in this thesis. On the other hand, the control subframe of the second channel may be used to speed up the completion of a TW handshake process. This may then assist in lowering the average end-to-end transmission delay of a given ladder network.

One approach is to further extend the TTDCA algorithm presented earlier in Section 6.2 to enable a grant IE message to be transmitted using the control subframe of the second channel. The procedures carried out within this enhanced TTDCA algorithm, hereafter referred to as the ETTDCA algorithm, are best described based on the example shown in Figure 6-3. Assume after a given node has sent a request IE using the second *transmission opportunity* in the control subframe of channel 1 (CH1), it sets its *temporary next transmission opportunity* (Tmp Next Xmt Opp) to be two transmission opportunities after the current one, i.e., transmission opportunity (Xmt Opp) No. 4. Note that, in the context of IEEE 802.16, Tmp Next Xmt Opp is referred to as the first transmission opportunity that a node may be able to compete with its neighbours for the right to send its next control message, such as the confirm IE. After this, the node will make use of the mesh election algorithm, as described in Section 3.3, to determine whether it can win a particular Tmp Next Xmt Opp.

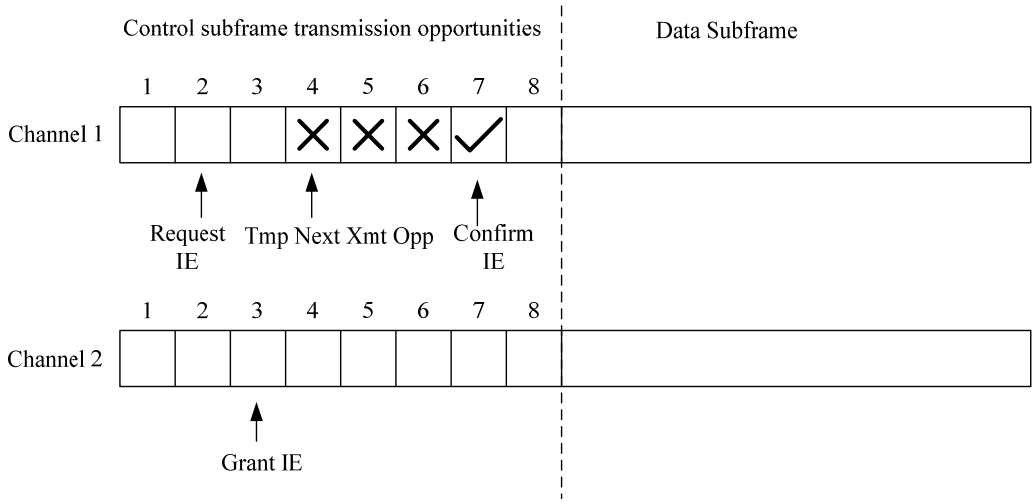



Figure 6-3 The ETTDCA algorithm allows the receiving node to send a grant IE on CH2 straight after it has received a request IE from a sending node on CH1.

The receiving node will transmit a grant IE back to the sending node in Xmt Opp 3 on channel 2 (CH2) using the second transceiver. The IEEE 802.16 standard specifies that with coordinated distributed scheduling, only one node within a two-hop neighbourhood is allowed to transmit a control message at a given time in a particular channel in order to avoid collision. In the example of Figure 6-3, only the sending node is transmitting its request IE in Xmt Opp 2 in CH 1. It is therefore possible for the node that is destined to receive this request to respond straight away by sending a grant IE using CH2 without the likelihood of a collision. This is because no other nodes will transmit on this channel at the particular instant.

Now, through the use of the second transceiver, a receiving node is able to respond to a request that it received in the other transceiver on CH1, by sending back a grant IE almost straightaway on CH2. In this way, a three-way handshake is very likely to be completed within a single frame, as depicted in the example of Figure 6-3. On the other hand, upon receiving the grant IE, the node may occasionally fail to win a transmission opportunity to send the confirm IE within the same frame on CH1. Even then, it is almost certain that the confirm IE could be sent the next frame as shown in Figure 6-4. This suggests that a given node will only be restrained from data transmission for at most one frame. As such, it is expected that the transmission

delay in a ladder network will be significantly reduced by the introduction of the ETTDCA algorithm.

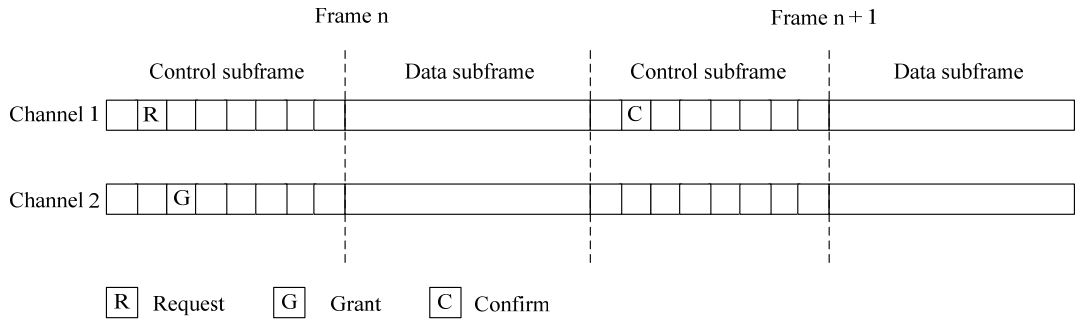



Figure 6-4 An example of a node being restrained from data transmission for one frame using the ETTDCA algorithm.

## 6.5 PERFORMANCE OF LADDER NETWORK OPERATING WITH ETTDCA ALGORITHM

### 6.5.1 Operating under normal condition

By replacing the TTDCA algorithm with the ETTDCA algorithm, the performance of the ladder network is again evaluated by means of computer simulation based on the same simulation settings as described in Section 4.4.1. Since the two algorithms differ only in the way a grant IE is handled during a TW handshake, but not in the resource allocation procedure for data transmission, the same request size of 110 minislots for a transmission link is again adopted here for the ETTDCA algorithm. As indicated in Appendix C.1, ladder networks of hop counts from 2 to 6 may make use of the same request size of 110 minislots. The simulated maximum achievable throughputs and average end-to-end packet transmission delays obtained for ladder networks of five different hop counts are tabulated in Table 6-4. Also shown are throughput and delay values achieved with four different *hexp* values. From Table 6-4, it is observed that the same maximum achievable throughput of 50 Mbps is achieved for all the four *hexp* values irrespective of the hop count of the ladder network. This is because the same request size of 110 minislots for a transmission link, as calculated in Appendix C.1, is applied to all the five ladder networks considered. As expected, the adoption of the ETTDCA algorithm only affects the average end-to-end delay. It is shown in Table 6-4 that for each hop count, a

minimum delay is obtained when *hexp* takes on a value of two, the same as for the TTDCA case. In fact the delays obtained with the ETTDCA algorithm at this *hexp* value are up to 18% lower than those presented in Table 6-3 for the TTDCA algorithm. Furthermore, the throughputs and delays obtained at this *hexp* value are similar to those shown in Table E-1 of Appendix E for the two parallel path networks. This implies that the frequency channel assignment plan, as shown in Figure 6-1(a), adopted in the ETTDCA algorithm, is effective in reducing the CDS of the ladder network to the same as that of the two parallel path network. Consequently, both networks are able to make use of the same request size to achieve the same throughput and delay.

Table 6-4 Maximum achievable throughputs and average end-to-end transmission delays obtained for ladder networks of five different hop counts and operating with four different *hexp* values using the ETTDCA algorithm.

| No of hops | Maximum achievable throughput (Mbps) | | | | Average end-to-end delay (ms) | | | |
|---|---|---|---|---|---|---|---|---|
| | *hexp* = 4 | *hexp* = 3 | *hexp* = 2 | *hexp* = 1 | *hexp* = 4 | *hexp* = 3 | *hexp* = 2 | *hexp* = 1 |
| 2 | 50 | 50 | 50 | 50 | 28.08 | 25.01 | 24.28 | 24.95 |
| 3 | 50 | 50 | 50 | 50 | 45.38 | 37.50 | 36.18 | 36.86 |
| 4 | 50 | 50 | 50 | 50 | 64.66 | 48.25 | 45.81 | 47.06 |
| 5 | 50 | 50 | 50 | 50 | 77.83 | 67.57 | 63.00 | 65.48 |
| 6 | 50 | 50 | 50 | 50 | 97.45 | 82.09 | 73.60 | 76.38 |

In general, packet transmissions in a ladder network of larger hop count will experience longer average end-to-end transmission delay because the traffic has to traverse a longer path before reaching the final destination. Moreover, the values of average end-to-end delay obtained with the ETTDCA algorithm, as shown in Table 6-4, are well below the upper limit normally allowed for services involving delay sensitive real-time traffic. This remains the case even for a six-hop ladder network. Also, although not shown here, it is worthwhile to comment that the throughput and average end-to-end delay achieved with both the ETTDCA and TTDCA algorithms follow the same trend, as shown in Figure 5-7 to Figure 5-12 for the single channel case, in terms of the buffer size used and the amount of traffic load applied.

### 6.5.2 Operating in the presence of a single node or link failure

When a ladder network suffers from a node or link failure, those links immediately before and after such a fault are required to handle the rerouted traffic in addition to their normal loads. For this reason, the request size for a transmission link not involved in handling any rerouted traffic will have to be adjusted downward. This is dealt with in greater detail in sections C.2 and C.3 of Appendix C for the case of a node failure and a link failure, respectively. The values of the request size used for evaluating the performance of the ETTDCA algorithm when applied to ladder networks of 2 to 6 hops are shown in Table C-2 and Table C-3 for the case of a node failure, and a link failure, respectively. For the computer simulations, the value of *hexp* is set at two, while all the other simulation settings are as described in Section 4.4.1. The simulated performance, in terms of maximum achievable throughput and average end-to-end packet transmission delay, for ladder network operating with the ETTDCA algorithm in the presence of a single node failure, is shown in Table 6-5.

Table 6-5 The maximum achievable throughputs and average transmission delays obtained for ladder networks of five different hop counts operating with the ETTDCA algorithm in the presence of a single node failure.

| Number of hops | Request size (minislots) | Maximum achievable throughput (Mbps) | Average end-to-end delay (ms) |
|---|---|---|---|
| 2 | 221 | 50.00 | 24.53 |
| 3 | 73 | 34.23 | 39.11 |
| 4 | | 34.23 | 52.77 |
| 5 | | 34.23 | 66.49 |
| 6 | | 34.23 | 79.76 |

As discussed in Appendix C.2, when a node failure occurs in a 2-hop ladder network, it becomes a single branch 2-hop chain network with only two links. For this special case, the two links operate on two separate frequency channels. As such, each of these two links is allocated the maximum request size of 221 minislots associated with a single frequency channel. On the other hand, a smaller request size of 73 minislots is applicable for ladder networks of 3 to 6 hops. The above explains why a 2-hop ladder network achieves a larger throughput of 50 Mbps, while those with 3 to

6 hops share the smaller throughput of 34.23 Mbps. The latter is 32% lower than the throughput obtained when the network is operating under normal condition. Moreover, the average delay is longer by at most 15% for the case of a single node failure. Nonetheless, these delays are well within the limit required for supporting real time traffics.

Performance evaluation is then repeated for the case when one of the links in a ladder network is faulty. In this case, the required request size for a surviving link not involved in handling rerouted traffic is calculated, in accordance to Appendix C.3, to be equal to 73 minislots. This value is applicable for all the five ladder networks considered. Note that, unlike the previous case with a single node failure, the 2-hop ladder network is now no longer a special case here. The simulated performance, in terms of maximum achievable throughput and average end-to-end packet transmission delay, for a ladder network operating with the ETTDCA algorithm in the presence of a single link failure is shown in Table 6-6.

Table 6-6 The maximum achievable throughputs and average transmission delays obtained for ladder networks of five different hop counts operating with the ETTDCA algorithm when any one of the links fails.

| Number of hops | Request size (minislots) | Maximum achievable throughput (Mbps) | Average end-to-end delay (ms) |
|---|---|---|---|
| 2 | 73 | 34.23 | 27.68 |
| 3 | | 34.23 | 40.54 |
| 4 | | 34.23 | 52.70 |
| 5 | | 34.23 | 67.54 |
| 6 | | 34.23 | 79.12 |

From Table 6-5 and Table 6-6, it is observed that with the exception of the 2-hop ladder network, the values for the maximum achievable throughput and average end-to-end transmission delay obtained are comparable whether a failure occurs in a link or node. It is interesting to note that the values obtained are not affected by the location of the fault. Moreover, when the values are compared with those obtained in the two parallel path networks, as shown in Table E-3 of Appendix E, it is observed

that the throughputs of the ladder network with hop count of three to six is 28% higher when compared to those obtained for the two parallel path network of the same hop counts. Also, the average end-to-end transmission delays of the ladder networks are slightly lower than that of the two parallel path networks.

## 6.6 PERFORMANCE OF THE ETTDCA ALGORITHM OPERATING WITH BIDIRECTIONAL TRAFFICS

As discussed in Section 5.6, in order to support bidirectional traffic transmissions, the total number of available minislots per frequency channel is divided into two equal sets, one for each direction. Also, with two-frequency two-transceiver operation, the two channels are distributed in the same way as described earlier in Section 6.2. Figure 6-5 shows the frequency plan of a 4-hop ladder network with the two frequency channels, CH1 and CH2, interchanged at alternate hop count levels. Also, the links at the same hop count level but transmitting in different directions are assigned with the same frequency channel. As for the nodes attached to a cross link, say node B and B', they will transmit via the link using the same frequency channel that they use to transmit in the forward direction, i.e., CH1. Since the total number of available minislots per frequency channel is being shared equally for transmissions in both directions, the request size for an individual link serving a given direction is now half of that used to support a unidirectional traffic flow.

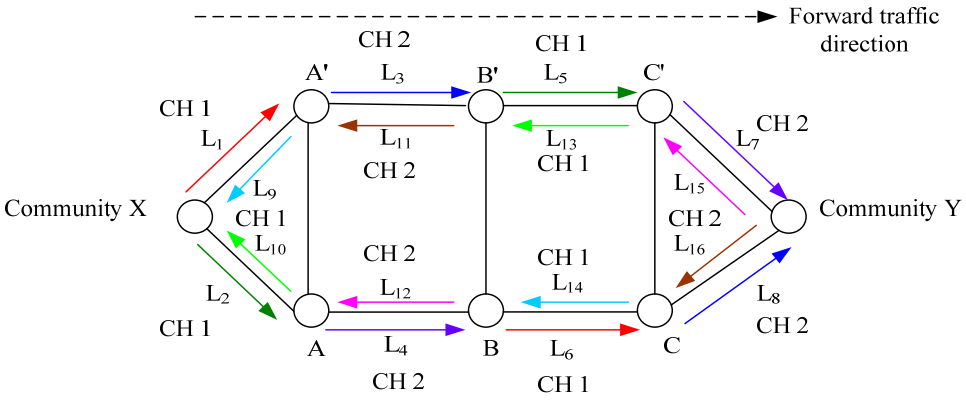



Figure 6-5 Channel allocations to individual links in a 4-hop ladder network. The nodes attached to a cross link will transmit via the link using the same frequency channel that they use to transmit in the forward direction.

For a ladder network of a given hop count, the request size for a transmission link is calculated in accordance to the procedure described in Appendix D. In the case of the network suffering from a node or path failure, the calculated request size is meant for a transmission link that is not involved in handling rerouted traffic. Note that the path between two adjacent nodes is made up of two links, one for each direction. In the case of a path failure, it is assumed that both of these links are down. The request sizes have been calculated for ladder networks with 2 to 6 hops. These are tabulated in Table D-1 for networks that operate normally, in Table D-2 for the case of a single node failure, and in Table D-3 when there is a path failure. Based on these values of request size, the performance of the ladder network operating with the ETTDCA algorithm in handling bidirectional traffics is evaluated by means of computer simulation. For the simulations, the value of *hexp* is set at two. The resultant maximum achievable throughputs and average transmission delays obtained for ladder networks of 2 to 6 hops are tabulated in Table 6-7 for operation under normal condition; in Table 6-8 for the case of a node failure, and in Table 6-9 for the case of a path failure.

Table 6-7 Maximum achievable throughputs and average transmission delays obtained for ladder networks of 2 to 6 hops operating normally with the ETTDCA algorithm in the presence of bidirectional traffics.

| No. of hops | Request size (minislots) | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 55 | 26.67 | 26.67 | 29.83 | 29.01 |
| 3 | | 26.67 | 26.67 | 46.87 | 46.04 |
| 4 | | 26.67 | 26.67 | 55.04 | 55.71 |
| 5 | | 26.67 | 26.67 | 72.22 | 71.78 |
| 6 | | 26.67 | 26.67 | 85.33 | 85.21 |

Table 6-8 Maximum achievable throughputs and average transmission delays achieved with the ETTDCA algorithm for bidirectional traffics in ladder networks of 2 to 6 hops in the presence of a node failure.

| No. of hops | Request size (minislots) | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 110 | 26.67 | 26.67 | 30.33 | 31.38 |
| 3 | 44 | 21.05 | 21.05 | 50.12 | 50.45 |
| 4 | 36 | 16.33 | 16.33 | 61.70 | 62.01 |
| 5 | | 16.33 | 16.33 | 79.55 | 78.63 |
| 6 | | 16.33 | 16.33 | 90.22 | 90.70 |

Table 6-9 Maximum achievable throughputs and average transmission delays achieved with the ETTDCA algorithm for bidirectional traffics in ladder networks of 2 to 6 hops operating when one of the intermediate transmission path fails.

| No. of hops | Request size (minislots) | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 44 | 21.05 | 21.05 | 35.94 | 36.65 |
| 3 | | 21.05 | 21.05 | 50.90 | 50.32 |
| 4 | 36 | 16.33 | 16.33 | 61.56 | 61.12 |
| 5 | | 16.33 | 16.33 | 78.89 | 78.33 |
| 6 | | 16.33 | 16.33 | 90.89 | 90.43 |

Table 6-7 shows that when a ladder network is operating with bidirectional traffics under normal condition, its throughput in any one direction is about 53% of that achieved for the case of unidirectional traffic. This is true irrespective of the hop counts of the ladder network. However, when the throughputs of both directions are added together, the total is slightly larger than the corresponding value shown in Table 6-4 for the case of unidirectional traffic. On the other hand, the average end-to-end transmission delay encountered by traffic in any one direction is at most 30% larger than what could be achieved with unidirectional traffic. Nonetheless, the

throughputs and delays obtained are the same as those tabulated in Table E-2 of Appendix E for the two parallel path network.

From Table 6-8 and Table 6-9, it is observed that the same maximum achievable throughput per direction of 16.33 Mbps is achieved for ladder networks with 4 to 6 hops, regardless of whether a failure occurs in one of the intermediate nodes or paths. This follows the same trend as the results of Table 6-5 and Table 6-6 for the case of unidirectional traffic. However, unlike the case of unidirectional traffic, the simulated throughputs for networks with lower hop counts of 2 and 3 are 30 % higher. Also, it is interesting to note that for each of the five networks considered, the total throughput of both directions is comparable with that obtained in the unidirectional case.

In the presence of bidirectional traffics, a ladder network equipped with the ETTDCA algorithm suffers an increase in average end-to-end transmission delay of less than 25 % during a node or path failure. Nonetheless, the actual delays are well within the limit normally specified for delay sensitive real time traffics, such as voice telephony and interactive multi-media services.

When comparing the above results with those obtained for the two parallel path networks in the presence of a failure, as tabulated in Table E-4, a lower throughput is obtained with the ladder network. This is due to the fact that a larger request size can be adopted in the two parallel path network when it becomes a simple chain network in the presence of a node or link failure. However, as discussed in Appendix E.2, such a network will cease operation when a failure occurs in both branches of the network. On the other hand, the ladder network is likely to survive such an event by rerouting traffic away from the failed nodes or links to arrive at the final destination.

## 6.7 SUMMARY

As discussed in the previous chapter, the use of the proposed RN and request-resend schemes has greatly mitigated the hidden node problem associated with the three-way (TW) handshake process for resource allocation in an IEEE 802.16 multi-hop

wireless backhaul based on a ladder network topology. Moreover, with single frequency operation, the number of available minislots to be shared among the various nodes and links is rather limited, particularly for networks with a large hop count. This greatly constrains the maximum throughput achievable for a ladder network. This problem becomes even more acute when the network is called upon to support bidirectional traffics.

In this chapter, a two-frequency two-transceiver distributed channel assignment (TTDCA) scheme is proposed for maximising the number of concurrent transmissions in order to obtain higher maximum achievable throughput in a ladder network. With TTDCA, each node is equipped with two time division duplex (TDD) radio transceivers to enable the node to simultaneously transmit and receive data, albeit on two different frequency channels, say CH1 and CH2. One of these channels is designated as the common channel during control subframe for adjacent nodes to exchange TW handshake messages. Through these control messages, a node will build up a record of the availability statues of minislots in a three-dimensional bit map, following the procedure described in [80, 81]. On the other hand, during intervals of the data subframe, both channels are used for data transmission and reception by each node in the network.

A simple frequency assignment scheme is proposed in Section 6.2 to enable the TTDCA algorithm to operate efficiently in a ladder network with two separate frequency channels. As shown in Figure 6-1(a), the two links at a given hop count level along the two parallel branches of a ladder network are pre-assigned to operate on the same channel. The two channels, CH1 and CH2 are then interchanged at alternate hop count levels. By operating the TTDCA algorithm in conjunction with the RN, request-resend and dynamic minislot assignment schemes, it is shown in Table 6-1 that there is more than two fold increase in maximum achievable throughputs compared with single frequency operation. This clearly suggests that the proposed TTDCA algorithm is effective in not only mitigating the hidden node problem but also enhancing the utilisation of available spectrum. Furthermore, the throughput remains constant for ladder networks with 2 to 6 hops. The ability to achieve a uniform throughput is attractive for network planning. In addition, there is a very significant improvement in the average end-to-end transmission delay. For

example, the average delay achieved by 6-hop ladder backhaul is slightly less than 90 ms, which falls well below the limit normally required for supporting real time services. This delay has been achieved by setting the value of the holdoff exponent (*hexp*) to two, which happens to yield the lowest delay among the four values of *hexp* considered.

With the TTDCA algorithm, only one frequency channel, i.e., the common channel, is utilised during intervals of the control subframe. Instead of leaving the second channel idle during a control subframe, an improvement to the TTDCA algorithm is proposed in Section 6.4. The new scheme is referred to as enhanced TTDCA or ETTDCA algorithm, which allows a receiving node to make use of the second channel to transmit its grant IE during a TW handshake process. In this way, a node is more likely to be able to complete a three-way handshake within a frame. This allows a 6-hop ladder network to achieve an 18% reduction in end-to-end transmission delay compared with the use of TTDCA algorithm.

The use of two frequency channels basically doubles the number of minislots available for distributing to individual nodes and links. This is even more significant when a failure occurs in a node or link. In such a situation, those nodes that have to handle the rerouted traffic would have to make use of a larger number of minislots to avoid traffic congestion. Computer simulations show that under the condition of a single node or link failure, it is possible to achieve a throughput in excess of 34 Mbps, a 32% reduction when compared with a ladder network that operates normally. Moreover, a node or link failure only leads to a 15% increase in average end-to-end delay. The same observations apply to ladder networks with a hop count of 3 to 6.

It is observed that the proposed ETTDCA algorithm is equally effective in supporting bidirectional traffics in ladder networks. In this case, the maximum achievable throughput in any one direction is about half that obtained with unidirectional traffic. In fact the total throughput from both directions becomes slightly more than that of the unidirectional case. The average end-to-end delays in a single direction are at most 30% larger than that for the case of unidirectional traffic.

Similar trends in throughput and delay are observed for ladder networks of different hop counts.

# CHAPTER 7

# CONCLUSIONS AND FUTURE WORK

## 7.1 CONCLUSIONS

Compared with conventional wired backhaul networks, wireless backhauls are potentially more cost effective, and easier to deploy, particularly in difficult geographical terrains. Also, a wireless backhaul network may be readily extended by adding extra nodes and links to create additional hops or branches to provide greater service coverage as the need arises in the future. This make a wireless backhaul network especially attractive for delivering low cost broadband services to less densely populated communities, often located in remote rural areas. Moreover, WiMax wireless technology, based on the IEEE 802.16 standard, has been successfully used for point-to-multipoint single-hop last mile broadband communications over a distance of up to 50 km. A mesh network version of this wireless technology has also been proposed for applications involving multi-hop transmission. The latter makes it an attractive candidate for use in wireless backhaul networks. To avoid disruption to services, it is necessary for a wireless backhaul to remain operational, albeit with a degraded performance, in the presence of node or link failures. As such, the main objective of this research has been the design of an IEEE 802.16 failure resilient wireless backhaul network capable of delivering broadband telecommunication services to a remote community from a regional or metropolitan centre. It is envisaged that the wireless backhaul will involve multi-hop transmission and meet the QoS requirement for delay sensitive services. Also, failure resilience is to be achieved with the use of minimum number of nodes. The main conclusions of this study are as follows:

- Chapter 2 provides an introduction to the conventional backhaul topologies, which include chain, tree, star and ring. Also, the various factors, such as, network deployment cost, type of failures, level of connectivity, traffic

- In Chapter 3, the frame structure and coordinated distributed scheduling for mesh mode operation, as specified in the IEEE 802.16 standard, are described in some detail. Moreover, the various means for improving the performance of the scheduling scheme, such as the alteration of holdoff exponent and holdoff base values, enhancement of data scheduling, as well as using multiple frequency channels for transmission, are examined. This reveals that there is still no effective scheme for alleviating the hidden node problem encountered in an IEEE 802.16 multi-hop wireless network.

- Following the literature review, a failure resilient backhaul network based on the IEEE 802.16 wireless technology has been proposed in Chapter 4. The proposed wireless backhaul takes the form of a ladder network, which connects two communities, separated by a long distance, using a minimum number of intermediate relaying nodes. There is a provision within the network of at least one backup path for each node pair, and this enables the backhaul to sustain multiple link and node failures.

- In Section 4.4.2, computer simulations, based on the NCTUns network simulator, have been carried out to evaluate the performance of the proposed ladder backhaul network operating with the IEEE 802.16 three-way (TW) handshake procedure incorporated into the coordinated distributed scheduling process. It is observed that for ladder networks of more than two hops, the maximum achievable throughput of the network degrades rapidly due to hidden node problem, which prevents the TW handshake to operate properly in allocating channel resources. As a result, a 92.7 % reduction in throughput is observed when the hop count of the network is increased from three to

- A novel reverse notification (RN) has been proposed in Section 4.4.4 in an attempt to alleviate the hidden node problem encountered during the IEEE 802.16 TW handshake. Computer simulated results, presented in Section 4.4.5, show that the use of RN has indeed improved the performance of the ladder network. In this case, the throughput achieved for a three-hop ladder network is now the same as its two-hop counterpart, which does not suffer from the hidden node problem. When compared with the original TW handshake, the introduction of the RN has allowed a 4-hop network to achieve a 22 times improvement in throughput, and a 69% reduction in average end-to-end transmission delay. In fact, this rather simple modification to the original TW handshake process has enabled ladder networks of 4 to 8 hops to achieve an end-to-end transmission delay of less than 150 ms, which is usually considered as the upper limit allowed for services involving delay sensitive real-time traffic.

- It is shown in Section 5.2 that the effectiveness of the proposed RN is reduced in the presence of a node failure, which tends to cause a RN message to arrive late at its destined node after it has already made a request. Based on this observation, a request-resend scheme is proposed to operate in conjunction with RN to more effectively overcome the hidden node problem encountered during a TW handshake. Furthermore, a dynamic minislot allocation scheme is proposed to adjust the number of minislots allocated to a node according to the amount of traffic load handled by the node during a node or link failure. The incorporation of these proposed schemes with RN in the TW handshake process has enabled ladder networks of two to six hops to achieve maximum throughputs that remain constant, regardless of the location of a node or link failure. Also, the average end-to-end transmission delay achieved for a 6-hop ladder network remains within the 150 ms limit required for supporting real time traffics. In fact, during normal operation, the throughputs of these ladder networks are approaching the theoretical maximum allowable traffic loads of the networks.

- In practice, a backhaul network is required to support bidirectional traffics. The performance of this mode of operation in the proposed failure resilient ladder network has also been evaluated by computer simulation. For single frequency operation, the maximum number of minislots available for resource allocation to individual nodes is fixed at 221 minislots, and this is divided equally for traffic transmission in each direction. As a result, the maximum achievable throughput of a ladder network in a given direction is expected to be half of that obtained when the network is supporting only unidirectional traffic. This has been verified by simulated results, presented in Section 5.6, for ladder networks of 2 to 6 hops operating normally or in the presence of a single node or link failure. Nonetheless, the total throughput from both directions is about the same as that of the unidirectional case.

- To enhance the maximum achievable throughput of a ladder network, it is necessary to increase the total number of minislots available for resource allocation. As such, a two-channel two-transceiver distributed channel assignment (TTDCA) algorithm has been proposed in Section 6.2 to support bidirectional traffic transmissions in a ladder network. Associated with this TTDCA scheme is a simple frequency assignment plan that assists in alleviating the hidden node problem, which would otherwise be made more troublesome by having to support communications in both directions. With the proposed TTDCA algorithm, it is able to double the maximum achievable throughput of a ladder network compared with single channel operation. This is true for ladder networks with hop counts of 2 to 6, and operating in either normal or failure conditions. Furthermore, the throughput achieved during normal operation remains constant regardless of the hop count of the network. In addition to the improved throughput, the average end-to-end packet transmission delays of the ladder networks are also significantly reduced. In particular, when operating with a holdoff exponent of two and a holdoff base of zero, it is possible to achieve a minimum transmission delay without sacrificing the maximum achievable throughput of the network.

- With the TTDCA algorithm, only one frequency channel is used during the control subframe while the other is left idle. A better resource utilisation is made possible by modifying the TTDCA scheme to make use of the second frequency channel to transmit grant IEs in a TW handshake. This new scheme is referred to as the enhanced TTDCA or ETTDCA algorithm. The introduction of ETTDCA has resulted in a reduction in average end-to-end transmission delay of up to 18%, compared with that achieved with the TTDCA scheme, while the maximum achievable throughput remains unchanged at 50 Mbps. When operating in the presence of either a single node or link failure, ladder networks of three to six hops are able to achieve the same throughput and average transmission delay. Most importantly, the transmission delay of a 6-hop network in the presence of bidirectional traffic is around 90ms, which is well below the 150 ms limit normally specified for real time traffics. Also, it is noted that under normal condition, the total throughput from both directions is slightly more than double that achieved with unidirectional traffic.

- The study has indicated that the TW handshake process of the IEEE 802.16 standard will need to be modified in order for it to operate satisfactorily in multi-hop transmissions. It is shown that the incorporation of the newly proposed ETTDCA algorithm into the IEEE 802.16 TW handshake process has enabled a ladder network to achieve high maximum throughput and low average transmission delay during both normal and failure conditions. This suggests that the proposed wireless ladder backhaul is not only able to support a great amount of data traffic, but is also suitable for relaying delay sensitive real time traffic, such as voice and video. Such a failure resilient wireless backhaul could form a cost effective option for providing reliable and high quality broadband communication services to remote rural communities.

- Although a ladder network has been adopted as the unique topology for in-depth study in this thesis, the proposed RN, request-resend, dynamic minislot allocation, TTDCA, and ETTDCA schemes are equally applicable to other

## 7.2 FUTURE WORK

The followings are recommended for future work.

- In this research, it is assumed that the geographical terrain between a remote community and a metropolitan centre is homogenous to simplify the design of a failure resilient wireless backhaul. This scenario is not likely in practice. Therefore, it becomes worthwhile to consider how the design of the network could be affected by the actual geographical terrains.

- The proposed failure resilient ladder network studied here consists of one traffic route linking a remote community and a metropolitan centre. It remains interesting and worthwhile to examine how the ladder network will perform when additional nodes and links are introduced to link with other remote communities.

- The use of directional antennas may assist in reducing co-channel interference among the nodes in a wireless network. It would therefore be beneficial to investigate whether its use could overcome some of the issues associated with the hidden node problem in ladder networks.

- This study concerns the use of traffic with constant bit rate (CBR) in an attempt to determine the maximum achievable throughput. It would be interesting to learn how well the backhaul network will cope with other types of traffic, including internet traffics, where the uplink and downlink rates are generally different.

# APPENDIX A

## CALCULATION OF REQUEST SIZE FOR A TRANSMISSION LINK IN THE PRESENCE OF A NODE OR LINK FAILURE

### A.1 CASE INVOLVING A NODE FAILURE

The request size of a transmission link in the presence of a node failure can be determined based on the same procedure described in Section 4.4.1. For example, with reference to Figure A-1 for a four-hop ladder network, when node C fails, link $L_4$ is identified to be associated with a largest collision domain set (CDS) of 7. Next, if we allow those links having the same colour to make use of the same set of minislots in a given frame, without the possibility of causing transmission collisions, then the largest CDS will be reduced from 7 to 4.

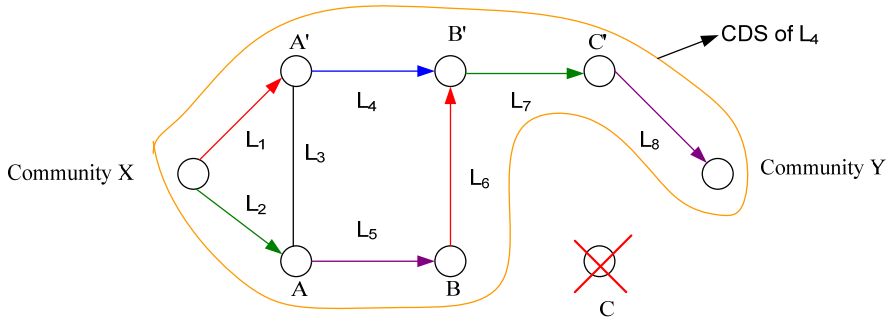



Figure A-1 Link $L_4$ is associated with the largest CDS of 7 when node C fails. Link $L_3$ is not included in the CDS as it is not used to handle rerouted traffic.

Now, unlike operating under normal conditions, some of the links in the network would need to handle a larger amount of rerouted traffic due to a node failure. Under such a condition, these links would need to have access to a larger number of minislots in order to avoid traffic congestion. When calculating the request size for a transmission link, it becomes necessary to first determine the number of links that are involved in handling the rerouted traffic. Again, referring to the example of Figure A-1, it is identified that link $L_7$, and $L_8$ will have to handle the additional rerouted traffic. Therefore, each of these two links will be allocated an extra set of minislots in

order to support the additional traffic rerouted to them due to failure of node C. To avoid possible transmission collisions, these extra sets of minislots must not overlap with those already in use by the other links. This means that the CDS value will have to be increased from four to six. In other words, the minimum number of required minislot sets becomes six. It follows that the maximum request size that could be allocated to each link, which is not involved in handling rerouted traffic, is given by

$$\text{Request size} = \left\lfloor \frac{\text{Total number of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor$$
$$= \left\lfloor \frac{221}{6} \right\rfloor = 36 \text{ minislots}$$

… (A.1)

where $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer.

Though the request size calculated above is determined for a failure occurring at node C, this value is equally applicable to a node failure at any of the other intermediate nodes in the four-hop ladder network. Computer simulations are used to verify that the use of the calculated request size can avoid the specific scenario depicted in Figure 4-13(a). Table A-1 tabulates the request sizes calculated according to the above procedure for ladder networks with two to six hops.

Table A-1 Request sizes allocated to a link not involved in handling rerouted traffic when a node fails in ladder networks having two to six hops.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 110 |
| 3 | 31 |
| 4 | 31 |
| 5 | 31 |
| 6 | 27 |

## A.2 CASE INVOLVING A LINK FAILURE

In the case of a link failure, the same procedure, as described in Section A.1, can also be used to calculate the request size of a link. With reference to Figure A-2, when the link $L_{11}$ in the four-hop ladder topology fails, link $L_4$ becomes associated with the largest CDS of 7. Again, if we allow the same coloured links to utilise the same minislots, the largest CDS is reduced from 7 to 4. Based on a CDS of 4, $L_8$ will have to handle double the amount of traffic compared with the other links. As such, it needs one additional set of minislots, which must not overlap with those already in use by the other links, in order to avoid transmission collision. This means that the CDS has to be increased from 4 to 5. Therefore, a minimum of 5 sets of minislots will be required to accommodate all the links. As such, the maximum request size for a transmission link, which is not involved in handling rerouted traffic, is given by

$$\text{Request size} = \left\lfloor \frac{\text{Total number of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor$$
$$= \left\lfloor \frac{221}{5} \right\rfloor = 44 \text{ minislots} \qquad \text{(A.2)}$$

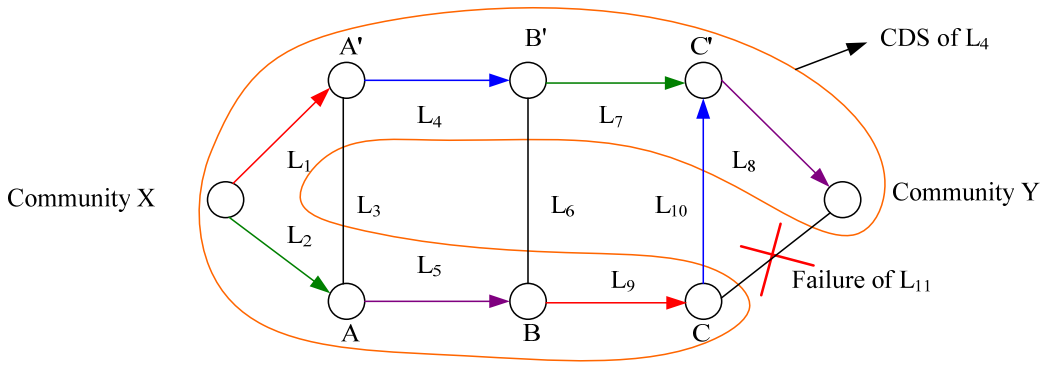



Figure A-2 Link $L_4$ is associated with the largest CDS of 7 when node C fails. Link $L_3$ and $L_6$ are not included in the CDS as they are not used to route traffic.

The above calculated request size remains valid for the failure of any link in the four-hop ladder topology. This calculated request size is then used as the initial value for computer simulations to find a more suitable value which could finally be adopted to avoid the undesirable situation depicted in Figure 4-13(a). Table A-2 shows the

request sizes obtained using the above described procedure for ladder networks of two to six hops.

Table A-2 Request sizes allocated to a transmission link not involved in handling rerouted traffic when a link failure occurs in ladder networks with two to six hops.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 44 |
| 3 | 36 |
| 4 | 36 |
| 5 | 36 |
| 6 | 31 |

# APPENDIX B

## REQUEST SIZE OF A TRANSMISSION LINK IN THE PRESENCE OF BIDIRECTIONAL TRAFFICS

### B.1 OPERATING UNDER NORMAL CONDITION

To support bidirectional traffic transmissions, each link will need to have access to two different sets of minislots, with one set for each direction as indicated by an arrow in Figure B-1. The procedure described in Section 4.4.1 is adopted here to identify the request size for each link in either direction. For example, according to Figure B-1, the link $L_3$ is associated with the largest collision domain set (CDS) of 12. Within this CDS, those same coloured links are able to make use of the same minislots for data transmission without causing collisions. Accordingly, the largest CDS may be reduced from 12 to 8. This also suggests that a minimum of 8 sets of minislots would be required to allow all links to transmit collision free.

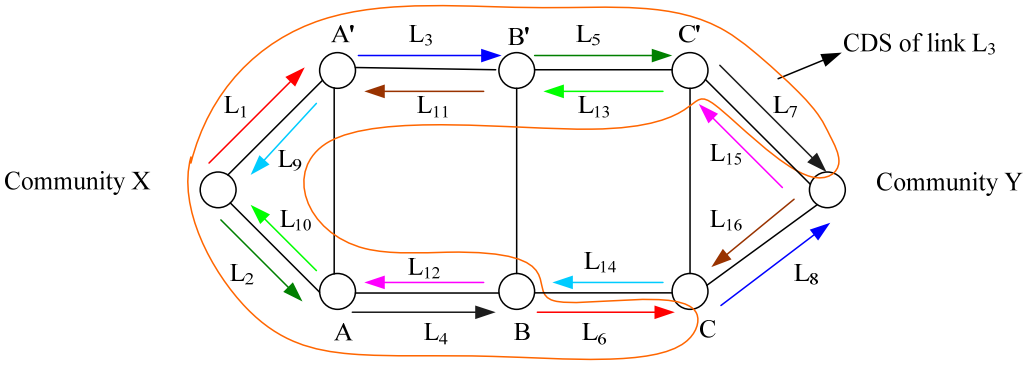



Figure B-1 Link $L_3$ in this four-hop ladder network is associated with the largest CDS.

It follows that the maximum request size for a link in a four-hop ladder network with bidirectional traffics is given by

$$\text{Request size} = \left\lfloor \frac{\text{Total number of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor = \left\lfloor \frac{221}{8} \right\rfloor = 27 \text{ minislots} \quad \text{(B.1)}$$

where $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer.

With the calculated request size used as the initial value, computer simulations are then carried out to verify its suitability for reducing the occurrence of the undesirable situation as depicted in Figure 4-13(a). Table B-1 shows the request sizes obtained for a link in ladder networks with a hop count from two to six.

Table B-1 Request sizes for a link in ladder networks with hop counts ranging from two to six.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 27 |
| 3 | 27 |
| 4 | 22 |
| 5 | 22 |
| 6 | 20 |

## B.2 CASE INVOLVING A NODE FAILURE

In a ladder network, when a node fails, the CDS value of a link will also change. It is therefore necessary to recalculate the CDS of each link in order to identify the largest CDS value. For example, when node C in the four-hop ladder network of Figure B-2 fails, it is identified that link $L_3$ becomes associated with a largest CDS value of 13. This CDS value is reduced to 8 if we consider that those same coloured links are able to make use of the same set of minislots for data transmission without causing transmission collisions. Next, we note that the three links, $L_6$, $L_7$, and $L_9$, are

involved in handling twice as much traffic as the other links. As such, these three links will require one additional set of minislots in order to be able to handle the traffic without congestion. However, the extra sets of minislots allocated must not overlap in time with those already used by the other links, in order to avoid possible transmission collisions. In this case, it means that we will need to increase the CDS value from 8 to 11. Hence, the maximum request size for a link, which is not involved in handling rerouted traffic, is calculated as

$$\text{Request size} = \left\lfloor \frac{\text{Total number of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor$$
$$= \left\lfloor \frac{221}{11} \right\rfloor = 20 \text{ minislots} \qquad \text{(B.2)}$$

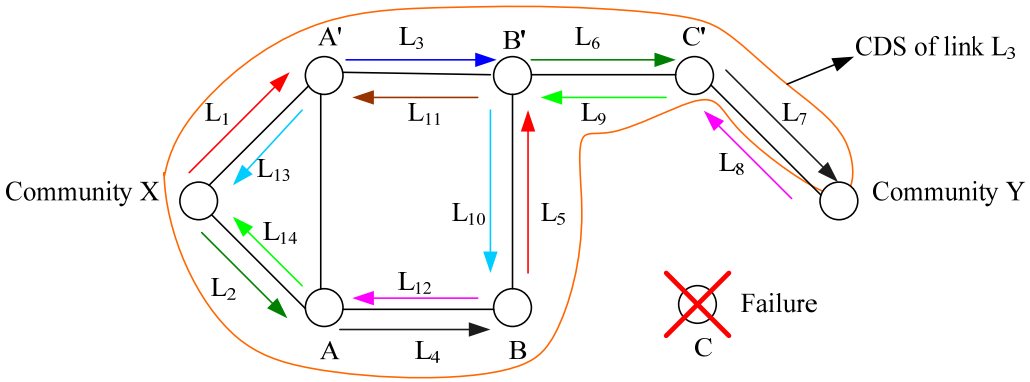



Figure B-2 Link $L_3$ is associated with the largest CDS when node C fails.

As the largest CDS value and the number of links involved in handling rerouted traffic remain the same, regardless of the location of a node failure, the calculated request size is applicable to a node failure occurring at any one of the intermediate locations. Once again, computer simulations are used to identify a more suitable request size, if available, for avoiding the undesirable situation as depicted in Figure 4-13(a). Table B-2 shows the request sizes for a link, which is not involved in dealing with rerouted traffic during a node failure, in ladder networks with hop counts of two to six.

Table B-2 Request size for a link, which is not involved in handling rerouted traffic during a node failure. For the two-hop ladder network, all the links are to handle rerouted traffic, and a request size of 55 is used for all links.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 55 |
| 3 | 20 |
| 4 | 18 |
| 5 | 18 |
| 6 | 17 |

## B.3 CASE INVOLVING A LINK FAILURE

The same procedure as described in the previous section is used to determine the request size of a link for a link failure occurring in a ladder network. Now, with reference to Figure B-3, the link $L_3$ is identified to be associated with a largest CDS value of 13 when the link between nodes C and Y fails. Again, if we allow all the same coloured links to make use of the same sets of minislots, then the largest CDS value will be reduced from 13 to 8. Within this reduced CDS, only one link, i.e., $L_7$, will be involved in relaying twice as much traffic compared with the other links. As such, $L_7$ will be allocated an extra set of minislots, which must not overlap with those minislots already in use by the other links, to avoid transmission collision. Taking this into consideration, the CDS value is then increased from 8 to 9. It follows that the maximum request size that can be allocated to a link not involving in handling rerouted traffic during a link failure is given by

$$\text{Request size} = \left\lfloor \frac{\text{Total number of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor = \left\lfloor \frac{221}{9} \right\rfloor = 24 \text{ minislots} \tag{B.3}$$

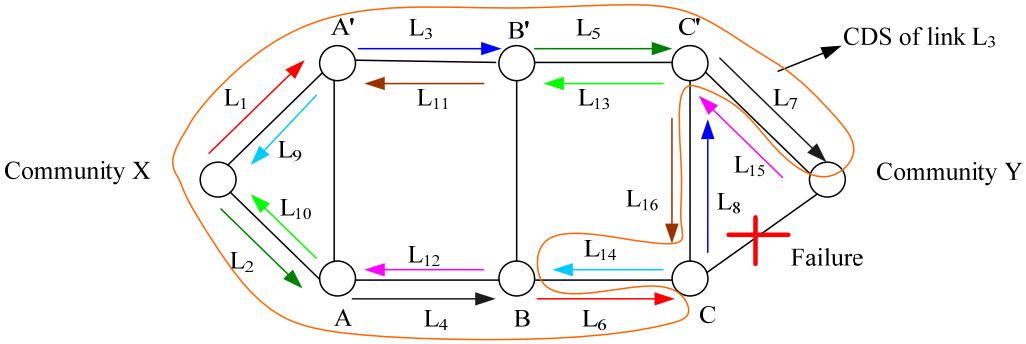



Figure B-3 Link $L_3$ is associated with the largest CDS during the failure of a link between nodes C and Y.

The request size calculated based on equation (B.3) is also applicable to any link not involved in handling rerouting during the failure of any intermediate link in a 4-hop ladder network. With the calculated request size used as the initial value, a more suitable value, if it is available, is then determined by computer simulation to overcome the problem depicted in Figure 4-13(a). The use of this refined request size is likely to lead to a better throughput for the network. Table B-3 tabulates the request sizes to be allocated to those links which are not involved in handling rerouted traffic during a link failure in ladder networks with 2 to 6 hops.

Table B-3 Request sizes for a link, which does not involve in handling rerouted traffic during a link failure.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 22 |
| 3 | 20 |
| 4 | 20 |
| 5 | 20 |
| 6 | 18 |

# APPENDIX C

## CALCULATION OF THE REQUEST SIZE FOR A TRANSMISSION LINK OF A LADDER NETWORK OPERATING WITH THE TTDCA ALGORITHM OR ETTDCA ALGORITHM

### C.1 OPERATING UNDER NORMAL CONDITION

When a ladder network is operating normally with either the TTDCA or ETTDCA algorithm, the request size for a transmission link can be determined following the same procedure as presented in Section 4.4.1. First, the link associated with the largest collision domain set (CDS) for a given frequency channel is identified. For example, for the four-hop ladder network of Figure C-1, link $L_1$, operating on CH1, is found to be associated with a largest CDS of three. Within this CDS, those links that share the same colour are allowed to make use of the same minislots without the likelihood of causing transmission collisions. In this case, the CDS value for $L_1$ is reduced from three to two. It follows that the request size for $L_1$ is calculated to be

$$\begin{aligned} \text{Request size} &= \left\lfloor \frac{\text{Total no. of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor \\ &= \left\lfloor \frac{221}{2} \right\rfloor = 110 \text{ minislots} \end{aligned} \tag{C.1}$$

where $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer.

Table C-1 shows the request size that could be allocated to a transmission link of a ladder network with a hop count from 2 to 6.

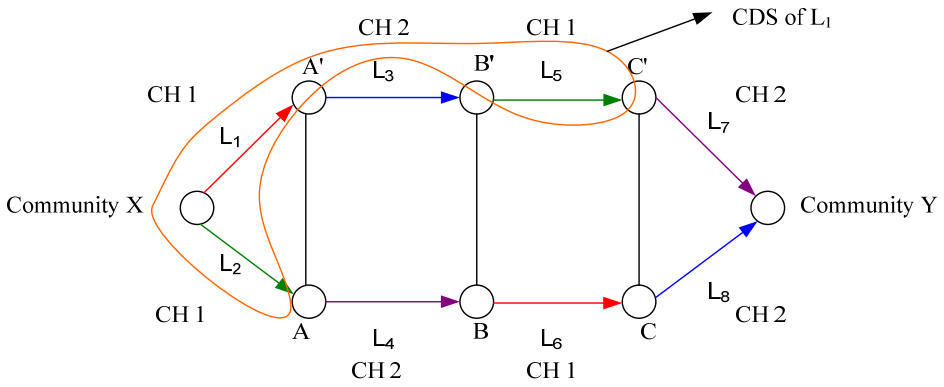



Figure C-1 $L_1$ is associated with the largest CDS.

Table C-1 Request size for a transmission link of a ladder network with a given hop count operating with either the TTDCA or ETTCDCA algorithm under normal condition.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 110 |
| 3 | 110 |
| 4 | 110 |
| 5 | 110 |
| 6 | 110 |

## C.2 CASE OF A NODE FAILURE

The procedure, as described in Appendix A.1, for calculating the request size for a transmission link in the presence of a node failure remains applicable for a ladder network operating with the ETTDCA algorithm. However, when it comes to identify the link associated with the largest CDS, it is only necessary to consider those links that share the same operating channel. For example, link $L_1$ in the four-hop ladder network of Figure C-2, has the largest CDS of three when node C fails. After accounting for the two same coloured links, $L_2$ and $L_7$, that are able to share the same set of minislots without the possibility of transmission collisions, the largest CDS is then reduced from three to two. Within this particular CDS, link $L_7$ is required to handle the extra rerouted traffic, and thus it is assigned one additional set of minislots in order to avoid traffic congestion. As a result, the CDS value is increased by one to

become three. It then follows that the request size that could be allocated to a link, which is not involved in rerouting traffic, is given by

$$\text{Request size} = \left\lfloor \frac{\text{Total no. of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor = \left\lfloor \frac{221}{3} \right\rfloor = 73 \text{ minislots} \tag{C.2}$$

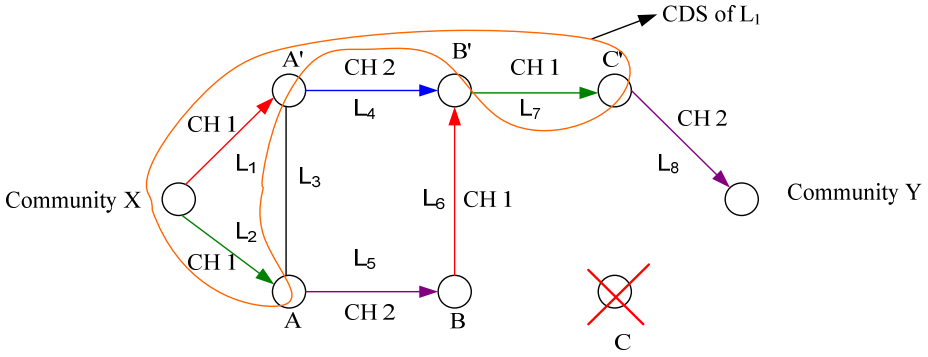



Figure C-2 A largest CDS of three associated with link $L_1$ when a failure occurs in node C. Note that link $L_3$ has not been considered as it is not used to route traffic.

It can be shown that the largest CDS value and the number of links that are needed for handling rerouted traffic remain the same regardless of the location of the failed node. As such, the request size calculated using equation (C.2) is applicable for any intermediate node of a given ladder network, with the exception of a 2-hop network. Table C-2 tabulates the request sizes calculated according to the above procedure for ladder networks with two to six hops operating with the ETTDCA algorithm.

Table C-2 The request sizes allocated to a link not involved in handling rerouted traffic when a node fails in ladder networks having two to six hops.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 221 |
| 3 | 73 |
| 4 | 73 |
| 5 | 73 |
| 6 | 73 |

Note that for a 2-hop ladder network, when a node fails, the remaining two operating links will have to handle the rerouted traffic. Since each of these two links is operating on a different frequency channel, it will therefore receive allocation of all the 221 minislots for that channel.

## C.3 CASE OF A LINK FAILURE

Again, consider the example of a 4-hop ladder network as shown in Figure C-3. If link $L_{11}$ fails, then we can follow the same procedure described in the previous section for a node failure, to calculate the request size for a transmission link not involved in handling the additional rerouted traffic. In this case, the request size calculated using equation (C.2) is applicable for any intermediate node of a given ladder network including a 2-hop network. Table C-3 tabulates the request sizes that could be used when a link failure occurs in ladder networks with 2 to 6 hops.

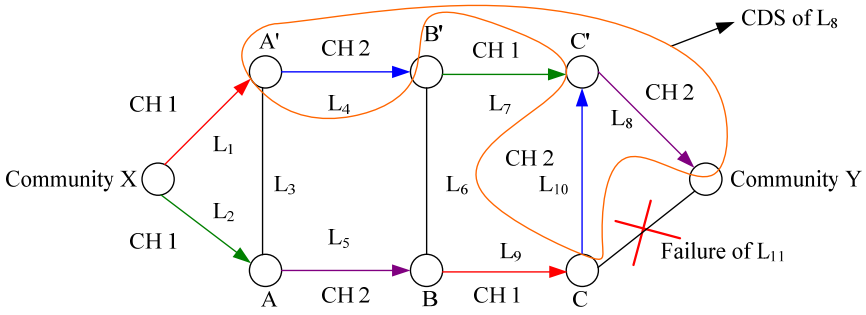



Figure C-3 A largest CDS of three associated with link $L_8$ when link $L_{11}$ fails. Link $L_3$ and $L_6$ are not taken into consideration as they are not used to route traffic.

Table C-3 The request sizes allocated to a link not involved in handling rerouted traffic, when a link failure occurs in ladder networks having two to six hops.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 73 |
| 3 | 73 |
| 4 | 73 |
| 5 | 73 |
| 6 | 73 |

# APPENDIX D

## CALCULATION OF REQUEST SIZE FOR A TRANSMISSION LINK OF A LADDER NETWORK OPERATING WITH THE TTDCA ALGORITHM OR ETTDCA ALGORITHM IN THE PRESENCE OF BIDIRECTIONAL TRAFFICS

### D.1 OPERATING UNDER NORMAL CONDITION

When operating with bidirectional traffics, each transmission path between two adjacent nodes is made up of two links, one for each direction. As the ETTDCA algorithm involves two transceivers operating on two different frequencies, the frequency allocation plan used is as shown in Figure 6.5. Again, the procedure presented in Appendix B.1 could be used to calculate the request size for a transmission link operating on a given frequency channel. For example, let consider a 4-hop ladder network as shown in Figure D-1. It is observed that for channel 1 (CH1), link $L_1$ is associated with the largest collision domain set (CDS) of 5. Within this CDS, those same coloured links could make use of the same set of minislots without the possibility of collisions. In view of this, the value of CDS could then be reduced from 5 to 4. Consequently, the request size for link $L_1$ is determined to be

$$\text{Request size} = \left\lfloor \frac{\text{Total number of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor$$
$$= \left\lfloor \frac{221}{4} \right\rfloor = 55 \text{ minislots} \qquad \ldots \text{(D.1)}$$

where $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer.

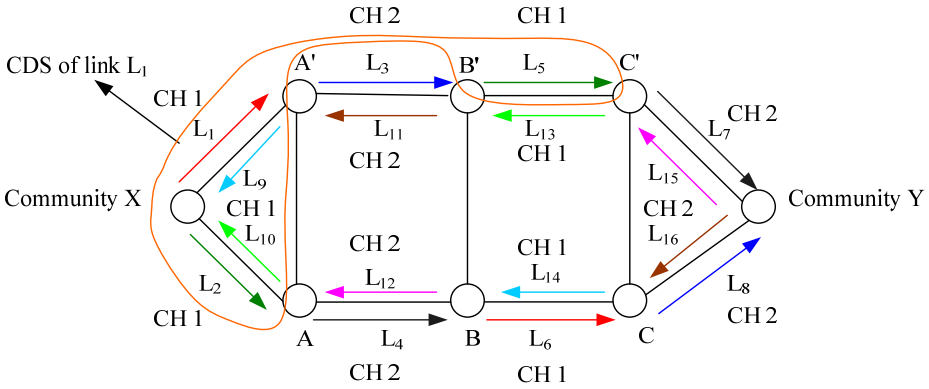



Figure D-1 Link $L_1$ is associated with the largest CDS of five.

The suitability of the calculated value for the request size, based on equation (D.1), could then be verified by means of computer simulation. In a similar way, the values of the request size that may be adopted for use in ladder networks with 2 to 6 hops are shown in Table D-1.

Table D-1 Values of request size for a transmission link in ladder networks with 2 to 6 hops.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 55 |
| 3 | 55 |
| 4 | 55 |
| 5 | 55 |
| 6 | 55 |

## D.2 CASE OF A NODE FAILURE

For a ladder network operating with the ETTDCA algorithm in the presence of a single node failure, the procedure described in Appendix B.2 for single frequency operation is still applicable for calculating the request size for a transmission link. In this case, the value of the largest CDS associated with a given link is based on a particular operating frequency. Now, if we again consider a 4-hop ladder network, as shown in Figure D-2, it is identified that link $L_5$ is associated with the largest CDS of 6 when node C fails. In view of the fact that links of the same colour within this CDS

may make use of the same set of minislots without the likelihood of collision, then the value of the CDS could be reduced from 6 to 4. However, within this CDS, links $L_6$ and $L_9$ are involved in relaying rerouted traffic. As such, each of these two links is allocated one additional set of minislots for this purpose. As these additional allocated minislots must not overlap in time with those minislots already assigned to other links, it is therefore necessary that the value of the largest CDS is increased from 4 to 6. It follows that the request size for a transmission link not involved in handling rerouted traffic in a four-hop ladder network is given by

$$\begin{aligned}\text{Request size} &= \left\lfloor \frac{\text{Total no. of data minislots in a frame}}{\text{Number of minislot sets required in the modified largest CDS}} \right\rfloor \\ &= \left\lfloor \frac{221}{6} \right\rfloor = 36 \text{ minislots}\end{aligned} \tag{D.2}$$

where $\lfloor \bullet \rfloor$ stands for rounding down to the nearest integer.

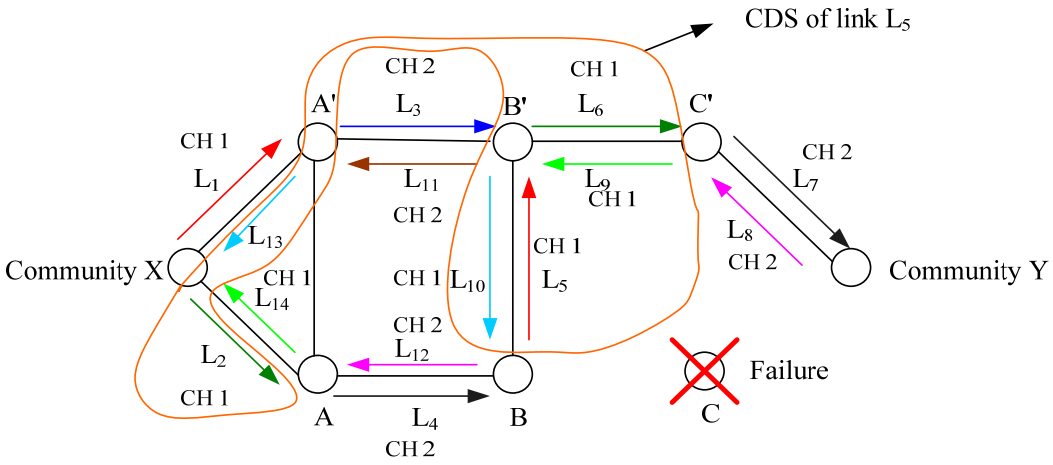



Figure D-2 Link $L_5$ is associated with the largest CDS of 6.

For ladder networks of 2 to 6 hops, the calculated request sizes are tabulated in Table D-2.

Table D-2 Values of the request size for a transmission link not involved in rerouting traffic in ladder networks of different hop counts operating with the ETTDCA algorithm in the presence of a node failure.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 110 |
| 3 | 44 |
| 4 | 36 |
| 5 | 36 |
| 6 | 36 |

Note that for a 2-hop ladder network, when an intermediate node fails, the remaining two operating links will have to handle the rerouted traffic. Since each of these two links is operating on a different frequency channel, it will receive an allocation of 110 minislots to transmit in one direction. As for the 3-hop ladder network, the largest CDS is identified at $L_1$ as shown in Figure D-3. The CDS is reduced to 4 after taking into consideration that the same coloured links can make use of the same minislots for data transmission. As only link $L_4$ is involved in rerouting traffic, the largest CDS is increased from 4 to 5. It follows that the request size for the links not involving in rerouting traffic is 44, as tabulated in Table D-2.

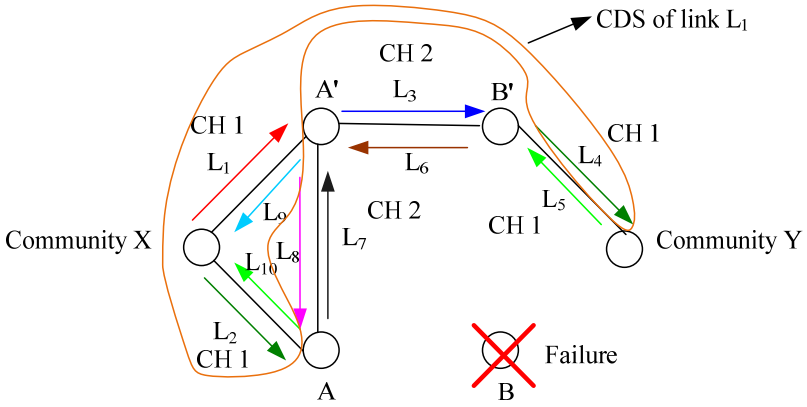



Figure D-3 Link $L_1$ is associated with the largest CDS of 5 when node B of 3-hop ladder network fails.

## D.3 CASE OF A LINK FAILURE

The procedure described in the previous section can also be adopted for determining the request size for a transmission link when a transmission path in a ladder network fails. Note that a transmission path between two nodes is made up of two separate links, one for each direction. When one of the transmission paths in the network, say the transmission path between nodes C and Y in Figure D-4, fails, it is observed that link $L_8$ is associated with the largest CDS of 6. After considering that the same coloured links within this CDS can make use of the same set of minislots without the possibility of collision, the value of CDS may be reduced from 6 to 4. However, since links $L_7$ and $L_{15}$ are required to relay rerouted traffic, each of them is allocated an additional set of minislots, which must not overlap in time with those minislots already in use by the other links. As such, the value of CDS will have to be increased from 4 back to 6. Based on this CDS, the request size for a transmission link can then be calculated using equation (D-2). Following the above procedure, the values of the request size for a transmission link, not involved in handling rerouted traffic, in ladder networks of five different hop counts are calculated and tabulated in Table D-3. Note that as the largest CDS for the 2-hop and 3-hop ladder networks are smaller than that of four-hop, five-hop, and six-hop counterparts, links in those networks are able to make use of a larger request size.

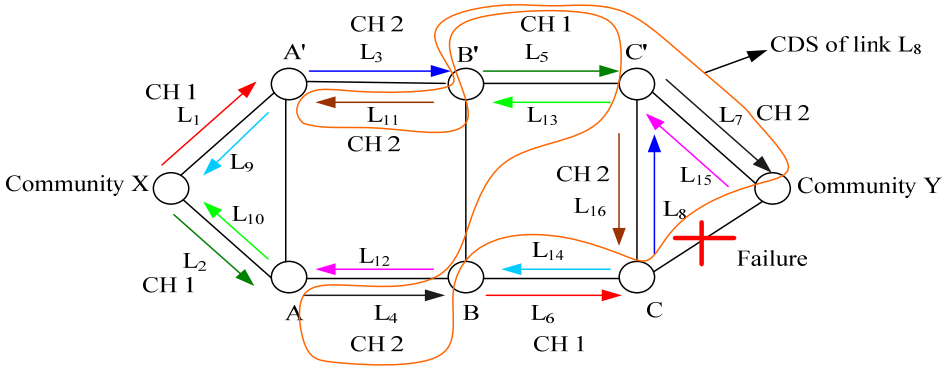



Figure D-4 Link $L_8$ is associated with the largest CDS of 6.

Table D-3 Values of the request size for a transmission link not involved in rerouting traffic in ladder networks of five different hop counts, operating with the ETTDCA algorithm in the presence of a transmission path failure.

| Number of hops | Request size (minislots) |
|---|---|
| 2 | 44 |
| 3 | 44 |
| 4 | 36 |
| 5 | 36 |
| 6 | 36 |

# APPENDIX E

## PERFORMANCE OF TWO PARALLEL PATH NETWORK OPERATING WITH ETTDCA ALGORITHM

### E.1 OPERATING UNDER NORMAL CONDITION

The performance of two parallel path networks of two to six hops operating normally with ETTDCA algorithm is evaluated by means of computer simulation based on the simulation settings described in Section 4.4.1. The performance evaluation is carried out under unidirectional and bidirectional traffic transmissions. The largest collision domain sets (CDS) of the two parallel path network operating under unidirectional and bidirectional traffic are the same as those of the ladder network. As such, the request sizes for a transmission link, i.e., 110 for the case of unidirectional traffic and 55 for servicing bidirectional traffic, are again adopted here for the two parallel path network. The resultant maximum achievable throughputs and average transmission delays obtained for two parallel path networks of five different hop counts under unidirectional and bidirectional traffic are tabulated in Table E-1, and Table E-2, respectively. As expected, with the use of same request size, the two parallel path network achieves the same throughput and delay as the ladder network of comparable hop count.

Table E-1 Maximum achievable throughputs and average end-to-end packet transmission delays of the two parallel path networks of five different hop counts in the presence of unidirectional traffic. The use of ETTDCA algorithm is assumed.

| Number of hops | Request size (minislots) | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 110 | 50 | 24.35 |
| 3 | | 50 | 36.54 |
| 4 | | 50 | 47.20 |
| 5 | | 50 | 64.71 |
| 6 | | 50 | 76.73 |

Table E-2 Maximum achievable throughputs and average transmission delays obtained for the two parallel path networks in the presence of bidirectional traffic. The use of ETTDCA algorithm is assumed.

| No. of hops | Request size (minislots) | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 55 | 26.67 | 26.67 | 30.36 | 30.75 |
| 3 | | 26.67 | 26.67 | 47.54 | 48.11 |
| 4 | | 26.67 | 26.67 | 56.32 | 55.42 |
| 5 | | 26.67 | 26.67 | 72.66 | 73.51 |
| 6 | | 26.67 | 26.67 | 86.90 | 85.35 |

## E.2 CASE OF A FAILURE

In the event of a node or link failure, the two parallel path network will become a single chain network. It is interesting to note that under such condition, the two parallel path networks of three to six hops have the same CDS value of 2, regardless of whether the network is relaying unidirectional or bidirectional traffic. This suggests that in either case, a transmission link will obtain 110 minislots for traffic transmission during a failure. As for a two parallel path network with the hop count of two, the two remaining links will operate on two different frequency channels, and each link will get 221 minislots when the network is handling unidirectional traffic, and half of the minislots for bidirectional traffic transmission. Based on these request

size values, the performance of the two parallel path networks of two to six hops, operating with the ETTDCA algorithm in handling unidirectional and bidirectional traffic, is evaluated using the same simulation settings as presented in Section 4.4.1. The maximum achievable throughputs and average end-to-end packet transmission delays obtained for the parallel path networks of 2 to 6 hops are tabulated in Table E-3 for the case of unidirectional traffic, and in Table E-4 for handling bidirectional traffic.

Table E-3 Maximum achievable throughputs and average end-to-end packet transmission delays of the two parallel path networks of five different hop counts operating under ETTDCA algorithm in the presence of unidirectional traffic during a node or link failure.

| Number of hops | Request size (minislots) | Throughput (Mbps) | Delay (ms) |
|---|---|---|---|
| 2 | 221 | 50.00 | 24.53 |
| 3 | 110 | 26.67 | 44.76 |
| 4 | | 26.67 | 56.47 |
| 5 | | 26.67 | 70.32 |
| 6 | | 26.67 | 85.13 |

Table E-4 Maximum achievable throughputs and average transmission delays of the two parallel path networks of two to six hops operating under ETTDCA algorithm in the presence of bidirectional traffic when any intermediate node or link fails.

| No. of hops | Request size (minislots) | Throughput (Mbps) | | Delay (ms) | |
|---|---|---|---|---|---|
| | | To community X | To community Y | To community X | To community Y |
| 2 | 110 | 26.67 | 26.67 | 30.33 | 31.38 |
| 3 | | 26.67 | 26.67 | 48.44 | 48.41 |
| 4 | | 26.67 | 26.67 | 57.19 | 56.78 |
| 5 | | 26.67 | 26.67 | 75.55 | 74.28 |
| 6 | | 26.67 | 26.67 | 87.69 | 86.98 |

The results shown in Table E-3 and Table E-4 have indeed verified that the two parallel path networks of three to six hops achieve the same throughput and delay

when handling unidirectional or bidirectional traffic. On the other hand, the two parallel path network with hop count of two, operating in the presence of unidirectional traffic, achieves almost double the throughput obtained under bidirectional traffic. When compared with the results obtained for the ladder networks of comparable hop count, as shown in Table 6-5, 6-6, 6-8, and 6-9, it is observed that the parallel path network, operating in the presence of unidirectional traffic, obtains a lower throughput even though the request size for a transmission link in the network is larger compared to that for the ladder network. This is due to the fact that the ladder network has two paths available for relaying traffic, each with 73 minislots. In the presence of bidirectional traffic, each link in the two parallel path network can obtain more than twice as many minislots as that obtained by a link in the ladder network. As a result, a higher throughput is obtained with the two parallel path network compared to the ladder network. However, the former will cease operation when a single node or link failure occurs in both branches of the network. On the other hand, the ladder network is likely to remain operational by rerouting traffic around the failed nodes or links via the cross links between the two branches.

# REFERENCES


[1] L. Ying-Chang, H. Anh Tuan, and C. Hsiao-Hwa, "Cognitive radio on TV bands: a new approach to provide wireless connectivity for rural areas," *IEEE Wireless Communications,* vol. 15, pp. 16-22, June 2008.

[2] O. Tipmongkolsilp, S. Zaghloul, and A. Jukan, "The Evolution of Cellular Backhaul Technologies: Current Issues and Future Trends," *IEEE Communications Surveys & Tutorials,* vol. 13, pp. 97-113, May 2010.

[3] T. Oishi, "VoIP and satellite systems," NSGDatacom 2005.

[4] R. Prasad and F. J. Velez, *WiMAX Networks: Techno-Economic Vision and Challenges*: Springer, 2010.

[5] A. P. Snow, U. Varshney, and A. D. Malloy, "Reliability and survivability of wireless and mobile networks," *Computer,* vol. 33, pp. 49-55, July 2000.

[6] B. Nleya and E. Nyakwende, "Survivability: Wavelength Recovery for Node and Link Failure in All Optical Networks," in *Third International Conference on Broadband Communications, Information Technology & Biomedical Applications*, Gauteng, South Africa, pp. 492-498, Nov 2008.

[7] K. Steiglitz, P. Weiner, and D. Kleitman, "The Design of Minimum-Cost Survivable Networks," *IEEE Transactions on Circuit Theory,* vol. 16, pp. 455-460, Nov 1969.

[8] C. L. Monma and D. F. Shallcross, "Methods for Designing Communications Networks with Certain Two-Connected Survivability Constraints," *Operations Research,* vol. 37, pp. 531-541, Jul 1989.

[9] L. W. Clarke and G. Anandalingam, "An integrated system for designing minimum cost survivable telecommunications networks," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,* vol. 26, pp. 856-862, Nov 1996.

[10] H. Frank and C. Wushow, "Connectivity considerations in the design of survivable networks," *IEEE Transactions on Circuit Theory,* vol. 17, pp. 486-490, Nov 1970.

[11] P. Leesutthipornchai, N. Wattanapongsakorn, and C. Charnsripinyo, "Efficient Design Techniques for Reliable Wireless Backhaul Networks," in *International Symposium on Communications and Information Technologies*, Lao, pp. 22-27, Oct 2008.

[12] C. Charnsripinyo and D. Tipper, "Topological design of 3G wireless backhaul networks for service assurance," in *5th International Workshop on Design of Reliable Communication Networks*, Italy, pp. 115-123, Oct 2005.

[13] C. Charnsripinyo and N. Wattanapongsakorn, "A model for reliable wireless access network topology design," in *IEEE Region 10 Conference*, Chiang Mai, Thailand, pp. 561-564, Nov 2004.

[14] H. Kawahigashi, Y. Terashima, N. Miyauchi, and T. Nakakawaji, "Designing fault tolerant ad hoc networks," in *IEEE Military Communications Conference*, Atlantic City, New Jersey, pp. 1360-1367, Oct 2005.

[15] S. Basagni, C. Petrioli, and R. Petroccia, "Fail-Safe Hierarchical Organization for Wireless Sensor Networks," in *IEEE Military Communications Conference*, Orlando, FL, USA, pp. 1-7, Oct 2007.

[16] H. Bin, T. Han, and X. Guoliang, "Fault-tolerant relay node placement in wireless sensor networks: formulation and approximation," in *Workshop on High Performance Switching and Routing*, Phoenix, Arizona, USA, pp. 246-250, April 2004.

[17] F. Kuhn, T. Moscibroda, and R. Wattenhofer, "Fault-Tolerant Clustering in Ad Hoc and Sensor Networks," in *26th IEEE International Conference on Distributed Computing Systems*, Lisboa, Portugal, pp. 68-77, July 2006.

[18] G. Gupta and M. Younis, "Fault-tolerant clustering of wireless sensor networks," in *IEEE Wireless Communications and Networking*, New Orleans, Louisiana, USA, pp. 1579-1584, March 2003.

[19] L. Yongxuan and C. Hong, "Energy-Efficient Fault-Tolerant Mechanism for Clustered Wireless Sensor Networks," in *Proceedings of 16th International Conference on Computer Communications and Networks*, Honolulu, HI, pp. 272-277, Aug 2007.

[20] G. Srivastava, P. Boustead, and J. Chicharo, "Link redundancy based connected topologies in ad-hoc networks," in *IEEE International Conference on Electro Information Technology*, Lincoln, Nebraska, pp. 6, May 2005.

[21] J. Owens, "Satellite backhaul viability," *Bechtel Telecommunication Technical Journal,* vol. 1, pp. 58-61, Dec 2002.

[22] T.-H. Wu, *Fiber Network Service Survivability*. Boston, MA, USA: Artech House, 1992.

[23] M. Grötschel, C. L. Monma, and M. Stoer, "Chapter 10 Design of survivable networks," in *Handbooks in Operations Research and Management Science*. vol. 7: Elsevier, 1995, pp. 617-672.

[24] Q. He, L. Cai, X. Shen, and P. Ho, "Improving TCP performance over wireless ad hoc networks with busy tone assisted scheme," *EURASIP Journal on Wireless Communications and Networking,* pp. 1-11, 2006.

[25] A. Konak and M. R. Bartolacci, "Designing survivable resilient networks: A stochastic hybrid genetic algorithm approach," *Omega,* vol. 35, pp. 645-658, Dec 2007.

[26] Y. Wu, J. Hui, and H. Sun, "Fast restoring gigabit wireless networks using a directional mesh architecture," *Computer Communications,* vol. 26, pp. 1957-1964, Nov 2003.

[27] M. Kodialam and T. V. Lakshman, "Dynamic routing of restorable bandwidth-guaranteed tunnels using aggregated network resource usage information," *Networking, IEEE/ACM Transactions on,* vol. 11, pp. 399-410, 2003.

[28] S. Ghosh, P. Ghosh, K. Basu, and S. K. Das, "GaMa: an evolutionary algorithmic approach for the design of mesh-based radio access networks," in *The IEEE Conference on Local Computer Networks*, Sydney, Australia, pp. 8 pp.-381, Nov 2005.

[29] M. Han, D. Fayek, and H. Pin-Han, "Availability-Constrained Multipath Protection in Backbone Networks with Double-Link Failure," in *IEEE International Conference on Communications*, Beijing, pp. 158-164, May 2008.

[30] B. T. Doshi, D. R. Jeske, N. Raman, and A. Sampath, "Reliability and capacity efficiency of restoration strategies for telecommunication networks," in *Fourth International Workshop on Design of Reliable Communication Networks*, Banff, Alberta, Canada, pp. 440-447, Oct 2003.

[31] M. Kodialam and T. V. Lakshman, "Dynamic routing of locally restorable bandwidth guaranteed tunnels using aggregated link usage information," in

*Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, Anchorage, AK, pp. 376-385 vol.1, April 2001.

[32] R. Doverspike and B. Wilson, "Comparison of capacity efficiency of DCS network restoration routing techniques," *Journal of Network and Systems Management,* vol. 2, pp. 95-123, 1994.

[33] S. Ramamurthy and B. Mukherjee, "Survivable WDM mesh networks. Part I-Protection," in *Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, New York, USA, pp. 744-751 vol.2, Mar 1999.

[34] Z. Yao, J. Fitchett, and K. Felske, "A real-time approach for fast failure restoration in dynamic optical networks," in *Fourth International Workshop on Design of Reliable Communication Networks*, Banff, Alberta, Canada, pp. 131-138, Oct 2003.

[35] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," *IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001),* pp. 0_1-857, 2004.

[36] J. Kamal, P. Jitendra, N. P. Venkata, and Q. Lili, "Impact of interference on multi-hop wireless network performance," in *Proceedings of the 9th annual international conference on Mobile computing and networking* San Diego, CA, USA: ACM, Sept 2003.

[37] G. A. Croes, "A Method for Solving Traveling-Salesman Problems," *Operations Research,* vol. 6, pp. 791-812, November 1, 1958 Nov 1958.

[38] S. Lin, "Computer solutions of the traveling salesman problem," *Bell System Techinical Journal,* vol. 44, pp. 2245-2269, Dec 1965.

[39] L. W. Clarke and G. Anandalingam, "A bootstrap heuristic for designing minimum cost survivable networks," *Computers & Operations Research,* vol. 22, pp. 921-934, Nov 1995.

[40] J. J. Shi and J. P. Fonseka, "Analysis and design of survivable telecommunications networks," *IEE Proceedings Communications,* vol. 144, pp. 322-330, Oct 1997.

[41] M. R. Wilson, "The quantitative impact of survivable network architectures on service availability," *IEEE Communications Magazine,* vol. 36, pp. 122-126, May 1998.

[42] A. Dutta and P. Kubat, "Design of partially survivable networks for cellular telecommunication systems," *European Journal of Operational Research,* vol. 118, pp. 52-64, Oct 1999.

[43] D. K. Pradhan and S. M. Reddy, "A Fault-Tolerant Communication Architecture for Distributed Systems," *IEEE Transactions on Computers,* vol. 31, pp. 863-870, Sept 1982.

[44] F. Houeto, S. Pierre, R. Beaubrun, and Y. Lemieux, "Reliability and cost evaluation of third-generation wireless access network topologies: a case study," *Reliability, IEEE Transactions on,* vol. 51, pp. 229-239, Jun 2002.

[45] M. K. S. Ho and K. W. Cheung, "Low Complexity Design of Fault-Tolerant Optical Network with Arbitrary Mesh Topology," in *Wireless and Optical Communications* Banff, Canada, July 2003.

[46] H. Kerivin, D. Nace, and T. T. L. Pham, "Design of capacitated survivable networks with a single Facility," *Networking, IEEE/ACM Transactions on,* vol. 13, pp. 248-261, April 2005.

[47] N. Garg, R. Simha, and X. Wenxun, "Algorithms for budget-constrained survivable topology design," in *IEEE International Conference on Communications*, pp. 2162-2166, April 2002.

[48] A. Frank, "Augmenting graphs to meet edge-connectivity requirements," in *31st Annual Symposium on Foundations of Computer Science*, St. Louis, USA, pp. 708-718, Oct 1990.

[49] I. Toshimasa and H. Masayuki, "Minimum augmentation of local edge-connectivity between vertices and vertex subsets in undirected graphs," *Discrete Applied Mathematics,* vol. 154, pp. 2307-2329, 2006.

[50] K. Jansen, et al., "Hardness of Approximation for Vertex-Connectivity Network-Design Problems," in *Approximation Algorithms for Combinatorial Optimization*. vol. 2462: Springer Berlin / Heidelberg, 2002, pp. 185-199.

[51] S. Wee-Seng, Z. Antoniou, and H. S. Kim, "Improving restorability in radio access network," in *IEEE Global Telecommunications Conference*, pp. 3493-3497 vol.6, Dec 2003.

[52] G. Gupta and M. Younis, "Fault-tolerant clustering of wireless sensor networks," in *IEEE Wireless Communications and Networking*, New Orleans, LA, USA, pp. 1579-1584 vol.3, March 2003.

[53] B. P. Crow, I. Widjaja, L. G. Kim, and P. T. Sakai, "IEEE 802.11 Wireless Local Area Networks," *IEEE Communications Magazine,* vol. 35, pp. 116-126, Sept 1997.

[54] V. Mohit, U. Shambhu, K. Vivek, and A. Vishal, "SAWAN: A Survivable Architecture for Wireless LANs," in *Proceedings of the Third IEEE International Workshop on Information Assurance* Maryland, USA, March 2005.

[55] F. E. de Deus, et al., "Fault tolerance in IEEE 802.11 WLANs," in *Telecommunications Symposium, 2006 International*, pp. 626-631, 2006.

[56] D. Qunfeng and Y. Bejerano, "Building Robust Nomadic Wireless Mesh Networks Using Directional Antennas," in *The 27th IEEE Conference on Computer Communications*, Phoenix, AZ, pp. 1624-1632, April 2008.

[57] Y. Bejerano and D. Qunfeng, "Distributed Construction of Fault Resilient High Capacity Wireless Networks with Bounded Node Degree," in *IEEE INFOCOM*, Rio de Janeiro, pp. 2646-2650, April 2009.

[58] G. Egeland and P. E. Engelstad, "The Economy of Redundancy in Wireless Multi-Hop Networks," in *IEEE Wireless Communications and Networking Conference*, Budapest, pp. 1-6, April 2009.

[59] "IEEE Standard for Information Technology-Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks-Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999),* pp. C1-1184, 2007.

[60] B. Sudeept, G. Samrat, and I. Rauf, "Design of IEEE 802.16-based multi-hop wireless backhaul networks," in *Proceedings of the 1st International Conference on Access Networks* Athens, Greece, Sept 2006.

[61] M. Cao, W. Ma, Q. Zhang, X. Wang, and W. Zhu, "Modelling and performance analysis of the distributed scheduler in IEEE 802.16 mesh mode," in *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, USA, pp. 78-89, May 2005.

[62] C. Cicconetti, A. Erta, L. Lenzini, and E. Mingozzi, "Performance evaluation of the mesh election procedure of ieee 802.16/wimax," in *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*, Chania, Crete Island, Greece, pp. 323-327, Oct 2007.

[63] C. Li-Der, L. Yen-Cheng, and C. Zi-Hong, "QoS coordinated distributed scheduling for 802.16 mesh networks," in *First Asian Himalayas International Conference on Internet*, Kathmandu, pp. 1-5, Nov 2009.

[64] Z. Ming, W. Suoping, and H. Tao, "Study on coordinated distributed scheduling in WiMAX mesh network," in *5th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1-4, Sept 2009.

[65] N. Bayer, X. Bangnan, V. Rakocevic, and J. Habermann, "Improving the performance of the distributed scheduler in IEEE 802.16 mesh networks," in *IEEE 65th Vehicular Technology Conference*, Dublin, pp. 1193-1197, April 2007.

[66] W. Shie-Yuan, L. Chih-Che, C. Han-Wei, H. Teng-Wei, and F. Ku-Han, "Improving the performances of distributed coordinated scheduling in IEEE 802.16 mesh networks," *IEEE Transactions on Vehicular Technology,* vol. 57, pp. 2531-2547, July 2008.

[67] V. Loscri, "A dynamic approach for setting parameters of the coordinated distributed scheduler of the IEEE 802.16," in *Proceedings of 17th International Conference on Computer Communications and Networks*, St. Thomas, US Virgin Islands, pp. 1-6, Aug. 2008.

[68] V. Loscri, "A queue based dynamic approach for the Coordinated distributed scheduler of the IEEE 802.16," in *IEEE Symposium on Computers and Communications*, Marrakech, pp. 423-428, July 2008.

[69] V. Loscri and G. Aloi, "Transmission hold-off time mitigation for IEEE 802.16 mesh networks: a dynamic approach," in *Wireless Telecommunications Symposium*, Pomona, CA, pp. 31-37, April 2008.

[70] L. Yun, F. SuiLi, Y. Wu, and L. HongZhi, "A dynamic approach for transmission holdoff time in IEEE 802.16 mesh networks," in *5th International Conference on Wireless Communications, Networking and Mobile Computing*, Beijing, China, pp. 1-4, Sept. 2009.

[71] K. Bong Chan, K. Dong Gu, S. Heecheol, L. Hwang Soo, and M. Joong Soo, "An adaptive holdoff algorithm based on node state for IEEE 802.16 mesh mode with coordinated distributed scheduling," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, Cannes, pp. 1-5, Sept 2008.

[72] C. de A Castro Cesar, N. L. S. da Fonseca, and S. V. de Carvalho, "Adjusting holdoff time of the IEEE 802.16 in mesh mode," in *3rd International Conference on New Technologies, Mobility and Security (NTMS)*, Cairo, pp. 1-5, Dec 2009.

[73] H. Bo, T. Fung Po, L. Lidong, and J. Weijia, "Performance evaluation of scheduling in IEEE 802.16 based wireless mesh networks," in *IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, Vancouver, pp. 789-794, Oct 2006.

[74] N. Bayer, D. Sivchenko, B. Xu, V. Rakocevic, and J. Habermann, "Transmission timing of signalling messages in IEEE 802.16 based mesh networks," in *12th European Wireless Conference 2006 - Enabling Technologies for Wireless Multimedia Communications*, Athens, Greece, pp. 2-5, Feb 2006.

[75] B. Makarevitch, "Distributed scheduling for WIMAX mesh network," in *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Helsinki, pp. 1-5, Sept 2006.

[76] B. Makarevitch, "Jamming resistant architecture for WiMAX mesh network," in *IEEE Military Communications Conference*, Washington, pp. 1-6, Oct 2006.

[77] T. Da, Y. Shoubao, H. Weiqing, and H. Yun, "TEOS: A throughput-efficiency optimal distributed data subframe scheduling scheme in WiMAX mesh networks," in *4th International Conference on Wireless Communications, Networking and Mobile Computing*, Dalian, pp. 1-4, Oct 2008.

[78] L. Hao-Min, C. Whai-En, and C. Han-Chieh, "A dynamic minislot allocation scheme based on IEEE 802.16 mesh mode," in *Second International Conference on Future Generation Communication and Networking*, Hainan Island, pp. 288-293, Dec 2008.

[79] W. Shie-Yuan, L. Chih-Che, and F. Ku-Han, "Improving the data scheduling efficiency of the IEEE 802.16(d) mesh network," in *IEEE Global Telecommunications Conference*, New Orleans, pp. 1-5, Nov 2008.

[80] C. Cicconetti, I. F. Akyildiz, and L. Lenzini, "Bandwidth balancing in multi-channel IEEE 802.16 wireless mesh networks," in *26th IEEE International*

*Conference on Computer Communications*, Anchorage, pp. 2108-2116, May 2007.

[81] C. Cicconetti, I. F. Akyildiz, and L. Lenzini, "FEBA: A bandwidth allocation algorithm for service differentiation in IEEE 802.16 mesh networks," *IEEE/ACM Transactions on Networking,* vol. 17, pp. 884-897, June 2009.

[82] S. Kuei-Ping, C. Hung-Chang, and C. Chi-Tao, "A decentralized minislot scheduling protocol (DMSP) for uplink and downlink traffic in IEEE 802.16 wireless mesh networks," in *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Tokyo, Japan, pp. 1143-1147, Sept 2009.

[83] L. Yu and Z. Zhang, "An enhanced mechanism to reduce the waste of allocated minislots in WiMAX mesh networks," in *International Symposium on Computer Network and Multimedia Technology*, Wuhan, pp. 1-4, Jan 2009.

[84] Z. Ming-Tuo, et al., "Multi-channel WiMAX mesh networking and its practice in sea," in *8th International Conference on ITS Telecommunications*, Phuket, pp. i-vi, Oct 2008.

[85] Y.-l. Tang, F. Yang, and Y. Yao, "A cross-layer scheme for multi-channel single-transceiver WiMax mesh networks," in *3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication*, Hong Kong, pp. 471-475, Aug 2009.

[86] D. Peng, J. Weijia, H. Liusheng, and L. Wenyan, "Centralized scheduling and channel assignment in multi-channel single-transceiver WiMax mesh network," in *IEEE Wireless Communications and Networking Conference*, Kowloon, pp. 1734-1739, March 2007.

[87] A. Ghiamatyoun, M. Nekoui, S. N. Esfahani, and M. Soltan, "Efficient routing tree construction algorithms for multi-channel WiMax networks," in *Proceedings of 16th International Conference on Computer Communications and Networks*, Honolulu, pp. 957-960, Aug 2007.

[88] C. Xiaoxuan, F. Xi, T. Xiaofeng, W. Yong, and Z. Ping, "Optimal deployment scheme for IEEE 802.16 mesh networks with combined single-radio and two-radio nodes," in *IEEE Vehicular Technology Conference*, Singapore, pp. 2854-2858, May 2008.

[89] A. Chengzhu and X. Jun, "The research of centralized scheduling algorithms in multi-channel multi-radio WiMAX mesh network," in *International Conference on Information Engineering and Computer Science*, Wuhan, pp. 1-4, Dec 2009.

[90] Y. Fan, T. Yu-Liang, C. Rung-Shiang, and T. Chen-Da, "Channel allocation for scheduling length minimization in WiMAX mesh networks," in *1st International Conference on Information Science and Engineering (ICISE)*, Nanjing, pp. 3989-3992, Dec 2009.

[91] A. Al-Hemyari, et al., "Cross-layer design using multi-channel system in WiMAX mesh networks," in *IEEE Region 10 Conference TENCON*, Hyderabad, pp. 1-6, Nov 2008.

[92] T. Yuliang, Y. Yan, and L. Xinrong, "A joint centralized scheduling and channel assignment scheme in WiMax mesh networks," in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*, Leipzig, Germany, pp. 552-556, June 2009.

[93] P. Mugen, L. Ming, and W. Wenbo, "A unified architecture and the tree level based scheduling technique for multi-radio multi-channel in IEEE 802.16 standard based wireless mesh networks," in *First International Conference on Communications and Networking in China*, Beijing, pp. 1-6, Oct 2006.

[94] M. Kas, B. Yargicoglu, I. Korpeoglu, and E. Karasan, "A survey on scheduling in IEEE 802.16 mesh mode," *IEEE Communications Surveys & Tutorials,* vol. 12, pp. 205-221, 2010.

[95] S. Y. Wang, et al., "The design and implementation of the NCTUns 1.0 network simulator," *The International Journal of Computer and Telecommunications Networking,* vol. 42, pp. 175-197, June 2003.

[96] Z. Hua and L. Kejie, "On the interference modeling issues for coordinated distributed scheduling in IEEE 802.16 mesh networks," in *Broadband Communications, Networks and Systems, 2006. BROADNETS 2006. 3rd International Conference on*, pp. 1-10, 2006.

[97] H. Zhu and K. Lu, "Performance of IEEE 802.16 Mesh Coordinated Distributed Scheduling under Realistic Non-Quasi-Interference Channel," in *International Conference on Wireless Networks*, Nevada, USA, pp. 379-385, Jun 2006.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.