NOMBRE DEL CURSO/CLAVE:	TE4017 Internet de las cosas y redes de sensores
NOMBRE DEL PROFESOR TITULAR:	Francisco Javier Ortiz Cerecedo
NOMBRE DEL LIBRO:	Guide to wireless mesh networks
NOMBRE DEL AUTOR:	Wang, J., Xie, B.
NOMBRE DEL CAPÍTULO:	Cap. 1 Journey from Mobile Ad Hoc Networks to Wireless Mesh Network
EDITORIAL:	Springer London
EDICIÓN:	1ed.
NÚMERO DE ISSN/ISBN:	9781848009097
FECHA DE PUBLICACIÓN:	2009
RANGO DE PÁGINAS:	1-30
TOTAL DE PÁGINAS:	30

Chapter 1
Journey from Mobile Ad Hoc Networks to Wireless Mesh Networks

Junfang Wang, Bin Xie, and Dharma P. Agrawal

Abstract A wireless mesh network (WMN) is a particular type of mobile ad hoc network (MANET), which aims to provide ubiquitous high bandwidth access for a large number of users. A pure MANET is dynamically formed by mobile devices without the requirement of any existing infrastructure or prior network configuration. Similar to MANETs, a WMN also has the ability of self-organization, self-discovering, self-healing, and self-configuration. However, a WMN is typically a collection of stationary mesh routers (MRs) with each employing multiple radios. Some MRs have wired connections and act as the Internet gateways (IGWs) to provide Internet connectivity for other MRs. These new features of WMNs over MANETs enable them to be a promising alternative for high broadband Internet access. In this chapter, we elaborate on the evolution from MANETs to WMNs and provide a comprehensive understanding of WMNs from theoretical aspects to practical protocols, while comparing it with MANETs. In particular, we focus on the following critical issues with respect to WMN deployment: *Network Capacity, Positioning Technique, Fairness Transmission* and *Multiradio Routing Protocols*. We end this chapter with some open problems and future directions in WMNs.

1.1 Introduction

A pure MANET is dynamically established by mobile devices grouped together as needed without any support from existing infrastructure as shown in Fig. 1.1. The mobile devices in the network communicate with each other through single or multi hop wireless links. The key benefit of MANET communication is that it enables

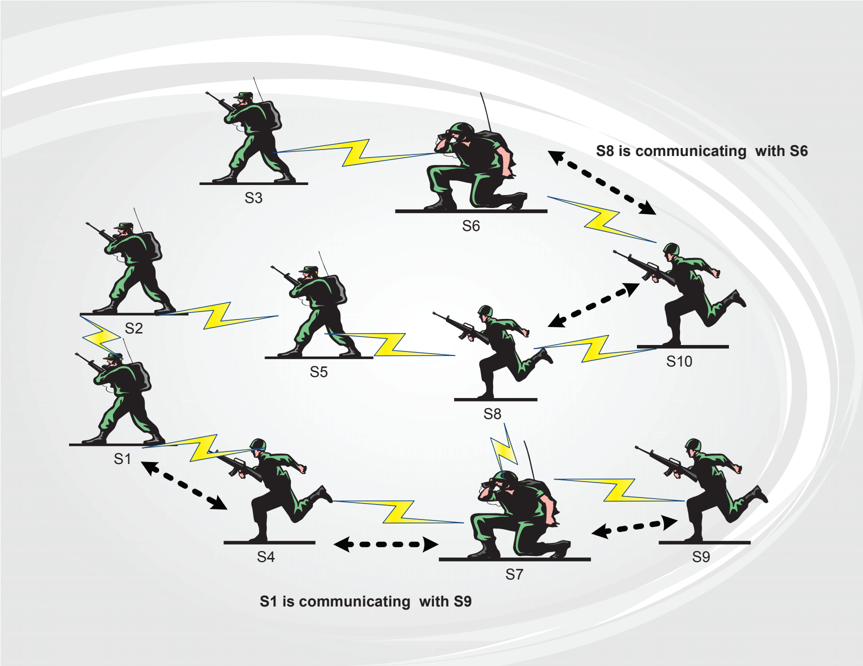


Fig. 1.1 An example MANET

us to form a network spontaneously without the need of having any infrastructure, which is both expensive and time-consuming.

The first MANET was initialized around 30 years ago by Defense Advanced Research Projects Agency (DARPA). Despite of peculiar advantages associated with MANETs, they have not been widely accepted for civilian applications. This could be primarily because of two reasons (1) some limitations of MANETs such as the security and limited throughput hinder MANETs from civilian applications and (2) the military and emergence applications dominate the research direction in MANETs so that most of the works target how to meet the unique requirements for these applications, such as the dynamic topology, which may not be solidly necessary for civilian applications.

In the recent years WMNs emerge as pragmatic multihop ad hoc networks to provide the high bandwidth Internet service to communities, enterprises, or entire cities. A WMN is a particular multihop ad hoc network, consisting of two parts: mesh backbone and mesh clients as shown in Fig. 1.2. The stationary wireless mesh routers (MRs) interconnecting through single/multi hop wireless links form the backbone. A few MRs with the wired connections act as the IGWs to exchange the traffic between the Internet and the WMN. The mesh clients can be the mobile wireless devices such as cell phones and laptops. The mobile clients connect to any MRs to access the Internet via the IGWs in a multihop fashion. Compared to pure MANETs, a WMN has a hierarchical structure and the topology of the wireless

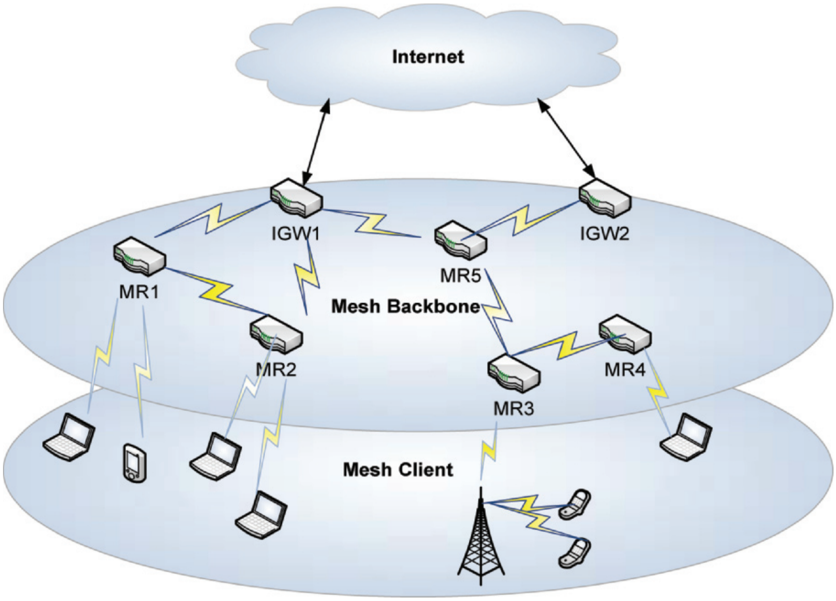


Fig. 1.2 A typical architecture of a WMN

backbone is relatively stable. These facilitate their deployment and application for Internet accessibility. Compared to the traditional wired and wireless networks (e.g., Wi-Fi), WMNs provide a cost-efficient way to support Internet services because of the reduced dependence on infrastructure by taking the advantage of multihop Internet access.

In this chapter, we first take an overview of the history of MANETs and WMNs. Then from theoretical aspects to practical protocols, we provide a comprehensive understanding of the evolution from MANETs to WMNs, concentrating specifically on the following issues:

- *Network capacity*. From the perspective of theoretical analysis, we present the capacity of MANETs (throughput per MR), and the capacity gain of WMNs in comparison with MANETs.
- *Positioning technique*. We detail the positioning technique of MRs and IGWs in WMNs that needs to satisfy various design constraints imposed by the system requirements.
- *Fairness transmission*. We explain the unfairness problem that exists in both MANETs and WMNs. We give special attention to end-to-end unfairness problem on WMNs compared to MANETs and provide an overview of the current approaches in WMNs.
- *Multiradio routing*. We study the routing issues in MANETs, including the design requirement and existing routing protocols. We also specify the requirements of routing protocols for WMNs in contrast to MANETs and then elaborate on some new routing metrics for WMNs.

1.2 Background: An Overview of MANETs and WMNs

In the past 30 years, MANET is one of the most active fields in communication networks, and has been experiencing tremendous advance in the past 10 years. In this section, we give an overview of MANETs and WMNs, covering the network architecture as well as the basic network characteristics.

1.2.1 Evolution of MANETs and Its Network Characteristics

The initial development of MANETs was primarily driven by military applications for the Department of Defense (DOD) of United States, in an attempt to provide a quick deployable communication system. In 1973, DARPA initialized the research on a new type of networks: Packet Radio Network (PRnet) [1]. The designed objective of PRnet was to allow communication among mobile devices by multihop wireless links. In 1983, DARPA started sponsoring Survivable Radio Network (SURAN) [2]. The goal of SURAN was to explore technologies that can enable the deployment of a large-scale packet radio network (i.e., up to 10,000 nodes). In 1994, DARPA launched the Global Mobile Information Systems (GloMo) [37] program, intending to develop new wireless ad hoc networking technology compatible with the Internet technology. At the same time, MANETs have received extensive attention from both the academia and industries because of explosive growth of personal wireless communication devices in the 1990s. In 1997, the Internet Engineering Task Force (IETF) established the Mobile Ad Hoc Network (MANET) Working Group to standardize MANET routing protocols such as Ad hoc On-Demand Distance Vector Routing (AODV) [3], Dynamic Source Routing (DSR) [4], etc.

The most prominent characteristic of a MANET is its infrastructureless. In other words, a MANET has no special centralized authority like cellular base station (BS) in support of wireless communication and is useful for a variety of communication applications where the infrastructure is unavailable for users. In brief, the main characteristics of MANETs can be summarized as follows:

- *Self-forming, self-configuring, and self-healing*. The multihop communication and its management in a MANET are automatic and spontaneous without any centralized network authority. In most scenarios, a MANET has no Internet accessibility for nodes.
- *Dynamic topology*. The locations of nodes in a MANET keep on changing because of node mobility and new node enrollment and leave without prior notification. Therefore, the network topology changes in an unpredictable manner.
- *Constrained resources*. MANETs suffer from limited energy and network bandwidth. The mobile nodes are usually powered by battery for transmitting or receiving packets to/from other nodes. Thus, the node can only work for a limited period because of limited power. In addition, all the nodes in a MANET usually operate in a shared wireless channel. Thus, the network bandwidth is limited and nodes compete with each other for accessing the medium.

The network performance (e.g., packet loss) of a MANET is significantly affected by path loss, interference, and other factors like other wireless networks. There are some new challenges in MANETs. For example, time-varying topology causes frequent link broken that needs path reconstruction to maintain the communication. The quality of service (QoS) of an application may be significantly impacted by link breakage and channel contention. Furthermore, the energy constraint requires the MANET design to be scalable for the purpose of energy-efficiency. Although MANETs offer the benefits such as self-forming, self-configuring, and self-healing so that each node is free to move while maintaining its communication, the challenging problems limit the application of MANETs. Currently, MANETs are still largely deployed for military networks rather than civilian applications.

1.2.2 Evolution from MANETs to WMNs

Since the late 1990s, the world has been experiencing both the prosperity of the Internet and the popularization of personal wireless devices. People always strive to design a low-cost and ubiquitous wireless environment that can allow their personal wireless devices to access Internet anywhere and anytime. WMNs are therefore proposed to satisfy the civilian application requirements while addressing the limitation of MANETs. In contrast to peer to peer structure of a MANET, a WMN consists of MRs forming a wireless backbone. As shown in Fig. 1.2, a number of MRs are interconnected by wireless links so that mobile devices are connected with the Internet via the wireless backbone. A MR not only handles the traffic from its associated mobile devices but also relays the traffic from other MRs. Similar to MANETs, a WMN still has the ability of self-configuring and self-healing, which keep some advantages of MANETs in a WMN. For example, the Internet access of a mobile device can be implemented in a multihop manner, which reduces installation cost of the overall system. The main improvements of WMNs over MANETs are as follows:

- *Infrastructure support of MANET*. In a MANET, the mobile device is usually not able to access the Internet. On the contrary, the IGWs of a WMN integrate with the Internet infrastructure, which provides the Internet accessibility for mobile clients that may not directly connected to the Internet. Internet accessibility is critical for civilian applications so that the Internet services such as email, Web browser, etc., can be enabled in WMNs. When the mobile device is moving in a WMN, it can handover its services from one MR to another. If it moves out the range of the WMN, the service continues if any other network is available. This may require an internetworking scheme for such mobility support.
- *Minimal mobility of MR*. In a WMN, each MR is situated in a fixed position, such as the roof of a building, and has minimal mobility. Therefore, the wireless backbone has a fixed topology unless some MRs or interfaces are added into or removed from the network. This is important to reduce the probability of link breakage and improve the network throughput.

- *Rich energy of MR*. In contrast to a MANET where the nodes are subject to energy constraint, every MR is usually connected to a rich power supply.
- *Multiradio and multichannel*. The network bandwidth is improved in WMNs because each MR can use multiple orthogonal channels. To manipulate multiple channels, a MR is configured with multiple radios (e.g., interfaces), which can simultaneously transmit/receive packets in multiple orthogonal channels without interference with each other. For example, the network capacity can be almost doubled if each MR has two radios to operate in two separate channels. As we discussed in the previous section, all nodes in a MANET only work on a shared channel with a single radio at each node. Table 1.1 lists the number of channels in IEEE 802.11s, the operating frequency, and their available data rate per channel.

These improvements with respect to the network architecture resolve some of the limitations and also improve the network performance as compared to the traditional wireless network such as MANET, cellular network, and WLAN. In terms of commercial implementation and application requirements, WMN offers many benefits and some of them are:

- *High-speed wireless system*. The wireless backbone of a WMN provides wireless accessibility for mobile users to communication with the Internet. IEEE 802.11g works in the 2.4-GHz band (like 802.11b) with a maximum raw data rate of 54 Mbps. With IEEE 802.11n products, the data rate between 100 and 200 Mbps becomes feasible for the mass market, which is not achievable in earlier wireless technology. The high data rate is enabled by high-speed Physical Layer (PHY) technologies such as sensitive Modulation Coding Schemes (MCS) and Multiple Input/Output (MIMO). Therefore, wireless backbone can be used to provide a platform for supporting various real-time commercial applications and bandwidth-consuming communication needs. The network capacity (throughput per MR) can be further improved by deploying multiple orthogonal channels over multiple interfaces. Meanwhile, the infrastructure nodes (i.e., IGWs) serve as the traffic sinks for the wireless backbone, and improve the network capacity when the number of IGWs asymptotically grows up with the number of MRs.
- *Promising coverage and connectivity*. The Internet accessibility is not available in a pure MANET because of its infrastructureless nature. In a WLAN, each access point is connected to the Internet and mobile devices are connected to the access point directly by using a centralized medium access. Therefore, the

Table 1.1 Multiple channels in IEEE 802.11

Protocol	Number of channels	Frequency (GHz)	Data rate (Mbps)	Transmission range (m)
IEEE 802.11a	12/13 indoor 4/5 outdoor	5	23–54 (2 stream)	30–100
IEEE 802.11b	14	2.4–2.5	5–11 (2 stream)	35–110
IEEE 802.11g	14	2.4–2.5	19–54 (2 stream)	35–110
IEEE 802.11h	12/13	5	≤100	30–100
IEEE 802.11n	14 in 2.4 GHz	2.4, 5	74–248 (2 stream)	70–160

coverage is largely limited and the connectivity is only available in a single hop range. A wireless personal area network (WPAN) provides high-speed data transmission, but it is only available for a short range because of limited transmission power and its operating frequency band. In a WMN, each MR not only operates as a host to aggregate data from its associated mobile clients, but also as a router to forward packets on behalf of other MRs. In this manner, the coverage and connectivity are significantly extended in a multihop fashion. Furthermore, the mobile device can fast and seamlessly handoff from an accessing MR to a neighboring MR at a faster rate by employing an appropriate migrating scheme.
- *Flexible and cost-efficient deployment*. WMN can be commercially operated in many ways. At first, it can be constructed and managed by a single Internet Service Provider (ISP). In such a case, each MR is under the control of an ISP. Secondly, a WMN can be semi-managed, meaning that the core MRs in a network are controlled by a single or several ISPs whereas the other MRs can be added by any users, subjecting to some operation and payment model. Moreover, a commercial WMN can be operated in unmanaged way that every MR is managed by an independent entity. For example, a user in a community can independently install a MR on the roof of its house and share the connection with a near-by ISP. Whichever commercial operating models, a MR is able to flexibly add or uninstall from a WMN because of self-configuration and self-forming. By taking advantage of IGWs, the Internet connectivity of a MR can be achieved by a multihop wireless connection. The loose integration of WMNs with the Internet significantly reduces the required wired links, which are expensive in link connection and hardware. Therefore, deployment of a WMN becomes a cost-efficient solution and is beneficial for mobile users.

Because of promising popularity in the market, several IEEE standard groups have been established to define specifications for WMN techniques in terms of different network types. In 2003, IEEE 802.15 and IEEE 802.11 each established a new subgroup, i.e., 802.15.5 and 802.11s, to standardize mesh network with their respective devices. IEEE 802.16a group and IEEE 802.20, the working group established in 2002 for mobile broadband wireless access networks, both include mesh into their standards as well. In the following subsection, we illustrate some example WMNs.

1.2.3 Free-For-Use and Commercial WMN Examples

Unlike MANETs, early stage of research works of WMNs have been performed on actual test beds or free-to-use networks. The first free-to-use WMN is Roofnet [5], which is an 802.11b/g community-oriented WMN developed by the Massachusetts Institute of Technology (MIT) to provide broadband Internet access to the users in Cambridge, MA. Some other free-to-use WMNs [6] have Champaign-Urbana Community Wireless Network (CUWiN), Seattle Wireless, Broadband and Wireless Network (BWN), and Southampton Open Wireless Network (SOWN), Technology For

All (TFA) [7]. These WMNs are typically implemented with open source software and free of addition of new nodes.

These test beds fostered tremendous advances and consequently built up the confidence for commercial applications in the design of architectures, protocols, algorithms, services, and applications of WMNs. Some of the commercial WMN solutions [8] include Tropos, BelAir, Cisco, Nortel, Microsoft, Firetide, Sensoria Corporation, PacketHop, MeshDynamics, and Radiant Networks. We illustrate them by using the example of Cisco.

Cisco [9] has commercial WMN solution by using its Aironet products to allow government, public safety, and transportation organization to build cost-effective outdoor WMNs for private or public use. The Aironet MR products are designed to provide secure, high-bandwidth, and scalable networks to enable access to fixed and mobile applications across metropolitan areas. The products use an 802.11 radio to provide network connectivity to the end users and a separate radio that is used for communication with the other MRs on the backbone. For instance, Aironet 1500 works as a MR that uses 802.11g/b for connecting the end users while using 802.11a radio to connect with neighboring MRs. The Aironet 1500 series support 16 broadcast identifiers to create multiple Wireless LANs so that the accessing network can be segmented to provide services to multiple user types. The IGW node is able to connect 32 other Cisco MRs (i.e., Aironet 1500).

A routing algorithm based on Adaptive Wireless Path Protocol (AWPP) allows remote MRs to dynamically select the multihop path toward the destination or the IGW. If new MRs are added to the network, each MR self-adjusts to ensure networking capability. The Cisco Aironet 1500 Series interacts with Cisco wireless LAN controllers and Cisco Wireless Control System (WCS) Software, having centralized key functions, scalable management, security, and mobility support. The security solution is compliant with 802.11i, Wi-Fi Protected Access (WPA2), and Wired Equivalent Privacy (WEP), which provide authentication for various WAP types and ensure data privacy with necessary encryption. The MR joins the network using X.509 digital certification and the wireless backbone uses the hardware-based Advanced Encryption Standard (AES) encryption. The Cisco solution based on the dual-radio approach raises the question of scalability and capacity in the infrastructure mesh, where all the MRs use 802.11a and the clients use 802.11b/g.

1.3 MANET and WMN Network Capacity

1.3.1 Capacity of Pure MANETs

Gupta and Kumar [10] derived the network capacity of a pure MANET. Intuitively, the network capacity is expected to increase with the network size because of spatial reuse of wireless channel. In fact, the theoretical analysis [10] has proved that the capacity of MANETs decreases with the growth of the network size. We consider a MANET with n MR nodes arbitrarily located in a disk R. Each node chooses an arbitrarily destination to which it wishes to send traffic at an arbitrary rate by single hop or multihops. If the channel has the capacity of w bits/s, the throughput per node in the MANET is $\Theta(w/\sqrt{n})$ bits/s [10]. In other words, the throughput per node decreases with the increase of the number of nodes. The main reasons of limited node capacity are:

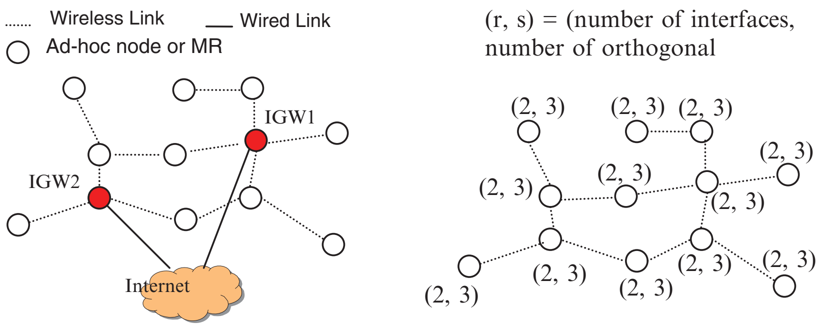
- *Multihop communication*. The traffic of a node has to be forwarded in a multihop fashion that repeatedly consumes radio resource.
- *Co-channel interference*. Concurrent transmission is not allowed by two nodes that are in the interference range. The wireless medium can only be spatially reused when the co-channel interference can be avoided.
- *Asynchronous sending and receiving*. The nodes usually cannot receive and send simultaneously at the same channel.

1.3.2 Capacity Gain of WMNs

In this part, we study the capacity gain of the WMN. Compared to MANET, WMN increases its network capacity in two ways: the addition of infrastructure nodes [11] and the usage of multiradio for operating multichannel [12].

1.3.2.1 Capacity Gain of Infrastructure

Liu et al. [11] studied the capacity gain by adding the infrastructure to a pure MANET, where access points like IGWs are placed either in a random or arbitrary manner. For instance, as shown in Fig. 1.3a, two IGWs are added in the MANET. It assumes that these IGWs are interconnected to the Internet with wired network of infinite bandwidth. Similar to an ad hoc node, each IGW has a data rate of w_1 bp over the common channel. The bandwidth w_1 is ivided into three parts: intracell w_1^1, uplink subchannel w_1^2, downlink subchannel w_1^3, where $w_1^2 = w_1^3$. Note that a



a MANET with infrastructure **b Multi channel and multi radio WMN**

Fig. 1.3 (**a**) MANET with infrastructure. (**b**) Multichannel and multiradio WMN deterministic routing scheme is used for a node to communicate. If the destination is outside the IGW in which the source node is located, the traffic is sent through the IGWs. Otherwise, the traffic is forwarded to the destination node hop by hop. Liu et al. [11] showed that the number of IGWs involved in multihop forwarding impacts the network capacity. If the number of IGWs (i.e., m) grows slower than $\sqrt{n}$, the network capacity per node scales as $\theta(w_1^1/\sqrt{n^* \log(n/m^2)})$. In this case, as a pure MANET, the network capacity decreases as the network size (i.e., n) increases. The improvement of network capacity because of IGWs is thus insignificant. On the contrary, if the IGWs are added at the rate faster than $\sqrt{n}$, the network capacity per node scales as $\theta(w_1^2\ {}^*m/n)$, meaning that the network capacity per node increases linearly as more IGWs are added. The IGWs reduce the number of wireless transmissions between nodes. Hence, IGWs can accommodate more traffic and effectively improve the network capacity. In order to achieve a nonnegligible capacity gain in an infrastructure-based MANET, the investment of IGWs should be high enough, i.e., the number of IGWs should grow at a rate faster than $\sqrt{n}$, as the number of ad hoc node increases. Thus, n node network with growing traffic requires a minimal increase in the number of IGWs such that the network capacity does not drop.

1.3.2.2 Capacity Gain of Multichannel and Multiradio

Kyasanur and Vaidya [12] investigated the gain of the network capacity because of multiply channels. They theoretically showed how multiple channels and interfaces can improve the network capacity. In a multichannel network, all nodes are equipped with multiple interfaces (r) to operate on s orthogonal channels, $1 \le r \le s$. In order to show the impact of multichannel, they consider a network without IGW nodes. As shown in Fig. 1.3b, each node has at most three wireless interfaces $(r = 3)$ to operate on three orthogonal channels $(s = 3)$. A MR is capable of concurrently transmitting or receiving traffic on multiple available interfaces at a given time. It assumes that the total data rate possible by simultaneously using all orthogonal channels is w_2. It is noted that $w_2 = s^*w$, if every channel has the data rate of w. If the number of channels is equal to the number of interfaces, all the channels can be fully utilized. Otherwise, depending on s/r, the network capacity may suffer capacity loss because of inefficient channel utilization. The analysis assumes that there is no delay in switching a channel from one interface to another. Otherwise, the network capacity will be further reduced, if no additional interfaces are provisioned at the nodes. The analytical results indicate that the maximal network capacity per node scales as $\theta(w_2/\sqrt{n^* \log n})$ bp when $1 \le s/r \le log\ n$ and the nodes are randomly located in the network domain. If $s/r \ge \log\ n$ (i.e., interfaces are less than channels than a rate), there is always a capacity loss. If the nodes can be arbitrary situated, the maximal network capacity per node scales as $\theta(w_2/\sqrt{n})$ bp with $r = s$. If $s > r$ (i.e., number of interfaces is less than the number of channels), the network capacity suffers a capacity loss. The above results reveal that the network capacity is improved in comparison with pure MANETs because of the use of multiple interfaces and multiple channels (i.e., $w_2 > w$). The actual improvement depends on the number of channels and interfaces, and the routing and transmission scheduling protocols used [12]. In order to achieve the maximal capacity improvement, all channels have to be effectively utilized, possibly by dynamically switching channels at the interface.

1.4 WMN Positioning Technique

As given in Sect. 1.2, the node in a MANET is randomly located and movable depending on the mobility pattern. On the contrary, the location of the MR in a WMN is stationary with minimal mobility. Again, the placement of the MRs and the IGWs is one of critical factors that determine the WMN performance [10]. If the MRs and the IGWs are randomly situated as in the MANET, it encounters the following problems:

- *Unbalanced load distribution*. The low traffic areas in a WMN domain may have many MRs whereas the heavy traffic areas could be covered with a few MRs. Thus, congestion may occur in the hotspots, resulting poor network performance in these areas. On the contrary, the MRs in the free traffic remain underutilized. Moreover, in a WMN, the traffic is primarily oriented towards the IGWs and the areas close to the IGWs have the high traffic. Hence, it is necessary to place more MRs or interfaces near the IGWs so as to alleviate congestion and achieve better network performance. The random deployment policy cannot take this factor into account and MRs around the IGWs may be the bottleneck of the traffic flow.
- *Uncontrolled interference*. Independent of its neighboring and global network information a random placement may create a network topology with high degree of interference, which could significantly deteriorate the network performance. For example, in an IEEE 802.11 wireless network, because of hidden terminal problem, the behavior of a node is decided not only on by its own capability, but also by the behavior of neighboring nodes and its hidden nodes. The achievable throughput for a MR decreases when the transmission rate of its neighboring nodes and hidden nodes increases. It is because of the fact that this MR may have a reduced chance to use channel/subchannel because of the presence of contention and interference. In contrast, a well positioned network can help in mitigating the interference by selecting the optimal position for each MR.
- *Unreliable architecture*. A random placement fails to consider the connectivity degree, which determines the fault-tolerance of the network. Therefore, the random placement approach results in an unreliable WMN architecture without considering fault-tolerance in the presence of link failures. It is possible that the failure of a link can disconnect the network so that all MRs, which connects to the IGWs by multi hop wireless links, may lose the Internet connectivity. For example, in a random constructed WMN, if an IGW fails, all MR connected to it may lose the Internet connectivity if there is no other neighboring IGWs that can provide an alternate connection.

Different locations of MRs and IGWs lead to different network topologies and architectures with distinct performance. Positioning technique in WMNs is required for appropriately placing the IGWs and MRs to achieve desired network performance. The WMN positioning technique can be defined as physical configuration of the IGWs and MRs including their locations and the number of interfaces on them. The configuration of IGWs and MRs is subjected to some constraints like geographical constraint, maximum number of channels, and the traffic demand. The positioning technique in WMNs can be further divided into two issues namely: IGW placement and MR placement as we study in detail.

1.4.1 Positioning Technique for IGWs

Positioning technique for IGWs targets to minimize the number of IGWs while meeting the bandwidth requirement. Because IGWs serve as the gateway to provide the Internet connectivity to MRs, IGWs usually are equipped with wired connections to the Internet, which increases the installation cost of a WMN. Therefore, to reduce the cost of the network, we need to keep the number of IGWs installed as small as possible. On the other hand, WMNs require enough IGWs to satisfy the network capacity needed by MRs. Therefore, the positioning technique for IGWs concerns where the IGWs should be located and how to minimize the number of IGWs while satisfying the MR Internet throughput demand. The critical questions should be answered while deploying IGWs:

- How many IGWs are needed in WMN?
- How many interfaces should be configured in the IGWs?
- Where the IGWs should be placed?
- How many and which MRs should be served by which IGW?

In the following section, we investigate Existing IGW selection algorithms that address these questions.

Cluster-based IGW selection. This deployment approach is proposed in [13], which is based on TDMA (Time division multiple access) MAC technology. In this scheme, the network nodes are divided into several disjoint clusters and in each cluster one IGW is deployed to serve the MRs in the cluster. To satisfy the minimal bandwidth requirement imposed by the QoS constraint, they assume each MR has a weight, which represents the bandwidth requirement. On the other hand, the total weight of all cluster nodes is also bounded because of limited capacity of the IGW. Then, the IGW deployment problem is abstracted as a clustering problem of minimizing the number of clusters that could satisfy the QoS constraints. Two polynomial time approximation clustering algorithms are proposed. The first algorithm, called shifting strategy, is a divide-and-conquer method and can have different computational complexity depending on different qualities of solutions. The shifting strategy is able to find a near optimal solution at the expense of high computational complexity. The second approach is a greedy domination-independent-set approach. This greedy algorithm selects a dominant-independent-set of the power graph as the set of cluster-headers, and guarantees an approximation ratio that is linear with the maximal radius *R* (i.e., the maximum number of hops from a MR to the IGW), independent of the network size. To satisfy the weight, depth and replay-load requirements, clusters created by either method is further refined based on the weight partition algorithm. To meet the assumption of the TDMA scheme, a special adaptive delivery mechanism is designed. Because of TDMA, this approach is not applicable for a generic WMN that is based on IEEE 802.11s.

Multihop WLAN-oriented IGW selection. The IGW positioning algorithm proposed in [14] is on the basis of IEEE 802.11 multihop WLAN network architecture. In their scheme, the constraints of wireless channel capacity, wireless interference, fault-tolerance, and variable traffic demands are all considered. With respect to co-channel interference, two coarse-grained interference models that capture the trend of throughput degradation because of wireless interference are proposed: the bounded hop-count model and smooth throughput degradation model. The IGW positioning problem is modeled as a capacitated facility location problem. There are two steps in their IGW positioning scheme: Given a set of potential IGW locations, they first prune the search space by grouping points into equivalent classes. Second, they use a greedy placement approach by which they iteratively pick an IGW that maximizes the total demands. In this step, the search on the IGWs is on the equivalence classes created in step one. To decrease the computing complexity, some approximations such as traffic demand are also used in the greedy IGW selection phase. In a WLAN, each node only has one interface and the drawback of this approach is that it fails to consider the issues of multichannel and multiradio.

OPEN/CLOSE Heuristic IGW selection. The problem of optimum IGW selection specifically for a mesh network has been analyzed in [15] recently. They have developed two approaches to select the optimal number of IGWs and determine the placement of IGWs. Given a network with MRs, the IGW selection approach determines the appropriate location for an IGW. The first approach is based on an integer linear programming model to greedily select IGWs from a set of potential alternatives and calculates all possible solutions of IGW placement in term of establishment cost and communication capacity between IGWs and MRs. But, the computational complexity of this approach increases with the number of potential IGWs, which limits its effectiveness for a large mesh network. The second approach is an OPEN/CLOSE heuristic to find a sub-optimal solution. The heuristic scheme starts from any feasible solution and repeatedly decreases the investment cost by a certain percentage. If no more solution can be found, the current solution is claimed to be the best approximation.

QoS-based IGW selection. A QoS-based IGW selection approach for WMNs in [16] developed a recursive algorithm with the purpose of minimizing the number of IGWs and satisfying the QoS requirement. In this approach, an one-hop dominating set of original graph will be greedily found first and this result will be used as the input of next recursion. The one-hop dominating set means that every node is connected to a clusterhead directly (i.e., single hop). The greedy dominating-set searching operation continue until the cluster radius reaches *R*, which is the predefined upper bound of cluster radius (i.e., maximum number of hop from a MR to the IGW). Finally, the clusterhead is the selected IGW. Because the cluster formulation fails to consider the hop length from each individual MR to the IGW, many faraway MRs may be attached to the clusterhead (i.e., IGW), and thus could not minimize the number of hop from the MR to the IGW.

Tree-based IGW selection. In [17], A WMN is modeled as a tree-based network architecture. In this architecture, the IGW is the root of the tree and all MRs are attached to the tree. The author first formulates the IGW selection as the problem of selecting multiple trees from an initial network graph by a linear program (LP). Then, they developed two heuristic algorithms: Degree-based Greedy Dominating Tree Set Partitioning (Degree-based GDTSP) and Weight-based Greedy Dominating Tree Set Partitioning (Weight-based GDTSP), for the purpose of cost-effective IGW selection. Both algorithms target to partition an initial network graph into multiple trees while considering the MR and IGW capacity. However, the tree-based WMN network architecture fails to consider the multipaths that allows each MR to connect the IGW using multiple paths.

1.4.2 Positioning Technique for MRs

The MR positioning problem can be described as a way to cover a given region with a minimum number of MRs and interfaces while meeting the traffic demand. According to this definition, the goal for MR positioning includes:

- Provide enough network capacity for satisfying the traffic demand
- Minimize the number of MRs and their configured interfaces, thereby minimizing the cost of deploying the network
- Avoid congestion incurred by balancing the traffic
- Maximize the network capacity with a certain number of MRs and their configured interfaces
- Provide fault-tolerance

The efficient placement of MRs is a challenging issue because of many practical constraints and contradictory requirements such as cost, link capacity, wireless interference, and varying traffic demands. For example, sparse MR placement is favored in terms of cost whereas dense placement provides better fault-tolerance. The main factors that should be considered in the MR placement are as follows:

- *Geographical restriction*. The MR placement should satisfy the coverage requirement expected by users and should also meet the location limitations imposed by the geographical terrain.
- *Connectivity degree*. The WMN consisting of MRs should be fully connected, meaning every MR is at least connected to an IGW by a path. On the other hand, a

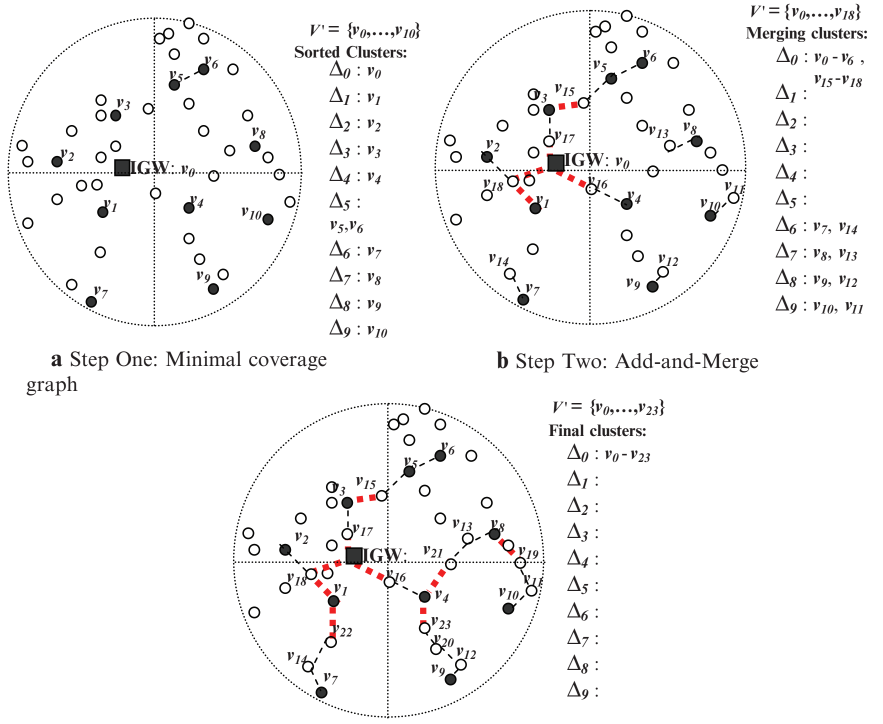
higher connectivity degree is beneficial to satisfy the variable traffic demand and is helpful in maintaining fault-tolerant links to provide resilient to MR failures.

- *Traffic information and link capacity*. The traffic information should be considered carefully for placing MR. In a heavy traffic area, more MRs or interfaces should be placed to meet the high traffic demand. The potential relaying path should be explored to avoid the congestion problem.
- *Co-channel interference*. The physical parameters of wireless link, such as fading and Signal-to-Noise Ratio (SNR), vary considerably in different geographical environments. The MR placement has to minimize the co-channel interference.

Compared to the research work on IGW placement, the study pertain to MR placement is still in the infancy time. The two network architectures commonly discussed in the most previous approaches are grid-based or tree-based [18]. In both these approaches, each MR has been assumed to be equipped with the same number of interfaces. In a grid-based mesh network, each node interface (e.g., 2, 3, or 4) connects to four neighboring nodes; whereas in a tree-based mesh network, each node connects to three neighboring nodes in which two nodes are descendent nodes in the tree. Both deployments are too simple for the real world. In addition, uniform interface configuration may again lead to poor performance and low utilization efficiency for the network architecture. For instance, in the tree-based architecture, a MR at a higher level needs to carry more traffic than its descendent nodes. They are supposed to require more radios. Otherwise, they might suffer from short of capacity. In contrast, those nodes far away from the IGWs may need fewer interfaces because they have less relaying traffic.

One of the elementary explorations about MR placement is presented in [19]. The authors discussed the MR placement on the condition that MRs can only be placed in the predecide candidate positions while considering the coverage, connectivity, and traffic demand constraints. They proposed a two-phase heuristic algorithm to find the optimal MRs and their positions. In phase I, the algorithm greedily excludes the candidate nodes that do not cause the uncovered hole by testing all the candidate nodes. The remaining node set, called coverage set, can satisfy the network coverage requirement but not the connectivity. For example, some nodes still have no route to the IGW. The nodes in the coverage set that can connect to each other form a cluster. In the other words, a cluster in the coverage set is the nodes they are connected at least by a path. In phase II, the algorithm adopted an add-and-merge procedure to select minimal number additional candidate nodes and add these nodes into clusters so that clusters can merge together. The mergence happened only if the aggregated traffic of the resulting cluster will not violate the traffic demand constraints.

Figure 1.4 illustrates an example how the algorithm proceeds. Initially, there are 43 candidate nodes as Fig. 1.4a shows. After removal of unnecessary nodes in phase I, it generates 11 nodes marked as black points and 10 corresponding clusters (i.e., Δ_i denotes cluster i as shown in Fig. 1.4a). It is because among 11 nodes, only nodes v_5 and v_6 are within the transmission range of each other and can be connected wirelessly. It can be seen that the nodes provide full coverage, but they still are not a connected graph. Figure 1.4b illustrates one step of add-and-merge procedure. Each of the dash lines represents that a new node has been added to a cluster. For example



a Step One: Minimal coverage graph

b Step Two: Add-and-Merge

c Step Three: Resulted network

Fig. 1.4 Heuristic algorithm for MR placement

v_{11} is introduced into Δ_9 and v_{15} into Δ_5. When a node is added, two clusters may be able to merge as one cluster. In Fig. 1.4, the square dot dash line represents the mergence of two connected-clusters. For instance, when v_{15} is added into Δ_5, Δ_5 is then able to merge with Δ_3 because v_{15} and v_3 are neighboring nodes. The final deployment showed in Fig. 1.4c includes 24 nodes after more nodes are added with a similar procedure.

1.5 Fairness Transmission in 802.11-Based MANETs and WMNs

IEEE 802.11 Distributed Coordination Function (DCF) is the most used MAC protocol in both MANETs and WMNs. DCF implements wireless medium sharing among a number of devices (i.e., nodes) through the use of carrier sense multiple access/collision avoidance (CSMA/CA) technology with a random back-off policy. A node first senses status of the channel for ongoing transmissions before sending a packet over a channel. If the channel is already in use, the node defers its transmission. It waits for a random time and re-attempts to sense the channel. On the other hand, if the channel is currently free, the node starts transmission. Such a mechanism is very effective when the channel is not overloaded, because it allows a node to transmit the packet immediately with a minimum delay. On the other hand, it always has a chance of collision if multiple nodes sense the channel free and begin transmission simultaneously. Although CSMA/CA solves the problem of channel contention, this MAC protocol, together with traditional transport protocols result in severe unfairness problems, which degrade the network performance:

- *Local unfairness* among the nodes that are within the interference range
- *End-to-end unfairness* between end-to-end multiple-hop flows

1.5.1 Local Unfairness

The phenomenon of local unfairness is that some flows dominate the transmission for a long period time while the other flows have no chance to seize the channel for their transmission. The main reasons that cause the problem are hidden and exposed terminal conditions and 802.11 MAC (i.e., CSMA/CA) backup policy.

If some devices within the interference range compete the medium for packet transmission, only one node can use the channel for transmission at a given time, while other node have to defer their transmission and enter into backup status as MAC protocol requires. Because the Transmission Control Protocol (TCP) also has the backup policy, the suspending transmission in MAC layer causes TCP to further back off. Thus, unfairness occurs and the losing flow suffers from a low transmission rate.

Figure 1.5 shows the two unfair node topologies: the hidden terminals and the exposed terminals. In Fig. 1.5a, S0 and S1 are hidden nodes and packets sent by the sender S0's can be corrupted by S1's signals. In other words, the reception at the receiver R0 fails because of S1's simultaneous transmission. On the contrary, packets sent from S1 to R1 are immune to S0's interference. Such a scenario causes unfairness because S1 has more chance for packet transmission than S0. In Fig. 1.5b, sender S2 can sense the other two senders but unfairness exists as following. In this

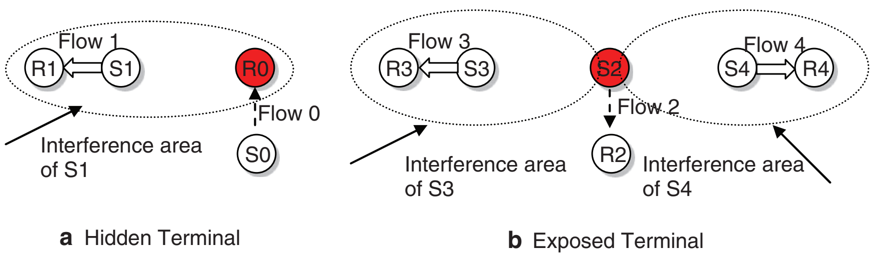


Fig. 1.5 Local unfairness

scenario, S2 is exposed to two senders (i.e., S3 and S4) whereas nodes S3 and S4 are only exposed to one sender (i.e., S2). Therefore, S2 is forced to back off more often, and thus flow 2 (S2−>R2) has less chance to transmit its packets than the other two flows. In order to evaluate the unfairness in the above scenarios, simulations are carried out for considering a case that the total bandwidth of a channel is 20 Mbps. The simulation results [20] have shown when the total bandwidth of a channel is 20 Mbps, Flow 0 and Flow 1 achieve 1 Mbps to 20 Mbps bandwidth allocation, respectively, in the scenario of Fig. 1.5a and the three flows in the scenario of Fig. 1.5b have the bandwidth of 18 Mbps, 3 Mbps, and 18 Mbps, respectively.

Local fairness has a similar impact on either 802.11-based MANETs or 802.11-based WMNs. For the purpose of providing fairness medium access in the 802.11-based network, some approaches have been proposed, and they can be classified into two categories:

General fair queuing. Some fair queuing algorithms are evaluated in [21]. The key point of wireless fair queuing algorithms is to monitor and predict the flow status and channel condition. The network then assigns more network resource to the node that experiences bursting and location-dependent errors in wireless channels, which compensates the losing flows. This approach requires a centralized coordinator such as access point that is able to schedule the wireless traffic in the network.

Self-adoption protocol. Self-adaptation in 802.11 networks has been recently addressed to provide a way for fair network bandwidth allocation. This approach includes altering the MAC back-off durations [22], switching from sender-initiate mode to receiver-initiate mode [23], and adapting the transmission rate and time scheduling [24]. For example, in the above hidden-terminal scenario in Fig. 1.5a, if the receiver R0 in Flow 0, rather than the sender S0, initiates the transmission by sending a request-to-receive (RTR) packet, then the sender S1 would back off while S0 is transmitting. Thus, Flow 0 no longer experiences packet losses [23]. On the other hand, in the exposed-terminal scenario, if the sender S2 shortens its random back-off window to a smaller value than the other two senders, then S2 will have a higher chance to use the shared medium than the other two senders [24]. These adaptation approaches, however, require modifications of the existing 802.11 MAC protocol.

1.5.2 End-To-End Unfairness

End-to-end unfairness is that a flow that has fewer hops achieves a high throughput than a long hop flow. There are three main reasons for this problem:

- Multiple-hop flows have to contend for the medium at each hop on its way to the destination node. Thus, flows that span more hops will spend more time on competing with more nodes in the network.

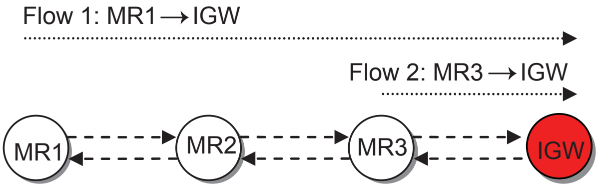


Fig. 1.6 End to end unfairness

- The packets forwarded on a longer path are more likely to be dropped than that of a shorter path.
- The end-to-end unfairness is getting even worse when the transport protocol, i.e., TCP, cannot distinguish the reasons of packet drop such as network congestion, packet collision, or stale route. Instead, it equally treats all packet losses as an indicator of network congestion and the sender will consequently reduce its congestion window size and lower the sending rate. These increases in round trip time (RTT) and packet loss lead to a lower throughput for the longer flows.

Figure 1.6 shows such an unfair example consisting of a three-hop TCP flow (i.e., Flow 1) and a one-hop TCP flow (i.e., Flow 2). Flow 1 and Flow 2 intend to transmit the same amount of packets. The experiments in [25,26] show Flow 1 always get far less share of the bandwidth than Flow 2 and starve most of time. The reason is that Flow 2 has an increased number of channel contention, an increased probability of packet loss, and more TCP backup times.

End-to-end unfairness problem in the WMN exhibits some unique characteristics compared to the MANET. The premise of MANET is that every node is interested in communicating with any other node in the network, and thus the flows could be well distributed in the network following different multihop paths. However, in a WMN, most of the packets are IGW-oriented, meaning that the packets are predominately transmitted between the MRs and the IGWs. In other words, the packets aggregated by a MR mostly lead to an IGW for the Internet, and the IGW sends the packets to the MR, upon receiving these packets from the Internet. Because of the fact that the flows originated by a distant node away now are more likely to share the same MRs with those flows closer to the IGWs, the queue delay accumulation will further slow the long-distance flows down. In other words, the default first in first out (FIFO) queue in each MR will make the packets of long-distance flows wait for a longer time while they are traveling a long distance toward the IGW. On the contrary, the traffic from the MRs closer to the IGWs can be forwarded to the IGW fast with a low packet loss. The following part is some approaches regarding to the end-to-end unfairness problem.

1.5.2.1 Global Fair Bandwidth Allocation

This approach requires a global view of the network topology and traffic load to make the decision on global fair network bandwidth allocation, exemplified by the

Inter-Transit Access Points (TAP) Fairness Algorithm (IFA) [25] and Co-ordinate Congestion Control algorithm (C3L) [26] protocols.

The IFA requires each MR first keep track of the original traffic (generated by its associated users) and the relay traffic separately. Then, each MR periodically broadcasts and forwards this statistical information to its neighboring MRs. Upon collecting this information from all the other MRs that reside on the same routing branch to an IGW, a MR is able to locally compute its own fair network share and correspondingly limit its sending rate according to the calculated result.

The C3L protocol is designed to achieve max–min per-flow fairness in the network. It first divides the whole network into several collision domains based on node topology; the network capacity of each domain is estimated by measuring the senders' queue length change while varying the traffic load pushed into that domain. It then allocates the network capacity to each node/link in that domain, proportionally to the number of flows going through it. The throughput of a multihop flow that crosses multiple collision domains will be determined by the smallest bandwidth granted by these domains.

1.5.2.2 Local Unfairness Measurement and Adjustment

Protocols such as ad hoc transport protocol (ATP) [27] and an end to end rate-based flow control scheme (EXACT) [28] only need local adjustments at each MR to help multihop flows to achieve the max-min fairness. These protocols first estimate the available bandwidth of each mesh link by measuring the queue, transmission, and contention delay at each MR. Then, each MR divides its bandwidth (or equivalently, the total service time) equally among all flows going through it. Thus, a multihop flow's throughput will be determined by the smallest bandwidth granted by the MRs along its routing path to the IGW. The difference between ATP and EXACT is that ATP calculates the "average packet service time" at each node, and the sender of each flow will adjust its transmission according to the maximum of these estimated service time. On the contrary, in EXACT each node divides the measured bandwidth among all the flows going through it.

1.6 Multihop and Multiradio Routing

1.6.1 Multihop Routing in MANETs

In a MANET, each node with a limited transmission range acts as traffic source, destination, or a router to collaboratively forward data packets for other nodes without a priori knowledge of the network topology. Therefore, a MANET routing protocol is required for a node to find the communication path in a dynamic network environment. In general, a routing protocol includes two parts: routing discovery and routing maintenance. In the routing discovery stage, the node discovers the path in a self-configurable manner before sending traffic. The basic idea of routing discovery is that the node initially advertises its existence and listens to the advertisement from its neighboring nodes. The node then knows the presence of its neighboring nodes as well as the way to reach them. The node can further find out the nodes that are outside its transmission range after its neighboring nodes have been discovered. In the stage of routing maintenance, a route can be reconstructed once the using path is broken. The routing reconstruction can be performed by a global or local routing discovery procedure to find the new path to the destination.

The design of a MANET routing protocol is a complex problem that has to consider many performance requirements (1) fast routing discovery, (2) fast routing recovery, (3) small communication overhead, (4) low computational overhead, and (5) efficiency and scalability for a large-scale network. These requirements are mostly imposed by the application requirements such as low traffic delay, node constraints such as limited power and memory, and node mobility. In the past few years, a variety of MANET routing protocols have been developed, and they can be classified into proactive, reactive, or hybrid, depending on the reaction of node in the routing determination process. On the other hand, they can be also categorized into flat or hierarchical protocols according to the logical network structure.

1.6.1.1 Proactive vs. Reactive

The proactive protocol maintains the routes for all destinations at every node up to date. Because every node always keeps the routes to all destinations, the communication connection can be immediately established any time by a node for reaching a destination, resulting in a minimal routing delay. In response to topology change because of node mobility, nodes have to periodically exchange the routing and link information between them even if the communication is not required. The proactive routing maintenance results in a heavy communication overhead in maintaining all routes up to date. Consequently, the network size and the node mobility are two hindrances in designing a scalable proactive routing protocol. The typical proactive protocols include Destination Sequenced Distance-Vector (DSDV) [29], Wireless Routing Protocol (WRP) [30], etc.

In contrast to proactive routing protocols, a reactive (or on-demand) routing protocol discovers or updates a multihop path from a source to the destination only when the communication is required. The multihop routing process is initiated only when a node has a packet to be delivered to a destination, and the path is maintained until the session is finished. Consequently, a reactive protocol avoids the prohibitive cost of routing maintenance as required in a proactive protocol, and thus achieves less communication overhead and scalability. On the other hand, it always suffers a delay on the path establishment before packet delivery if there is no fresh route to the destination. The typical reactive routing protocols include DSR [4], AODV [29], etc. The detailed information can again be found in [31].

1.6.1.2 Flat vs. Hierarchical

In a flat routing protocol, the MANET does not need to maintain any specific structure such as hierarchy, and every node performs the same functionalities. Upon receiving a routing request from a node, all neighboring nodes response by forwarding it to its neighboring nodes if itself is not the destination node. The flat protocol saves the communication overhead caused by maintaining a hierarchical network structure. However, the scalability and complexity may be problems in a large-scale network. For example, AODV is a flat routing protocol.

Rather than a flat network structure, a hierarchical routing protocol maintains a hierarchical network structure, which offers scalability and reduces the complexity in routing computation as well. In general, a hierarchical routing protocol organizes the network as a hierarchy, consisting of certain number of clusters, from the lowest level 0 to the top level $L-1$. In each level, it may have multiple clusters and each cluster elects certain nodes as the leaders. The cluster leaders maintain network state information at multiple levels of granularity. A level i $(0 \leq i < L)$ cluster consists of level $(i+1)$ clusters. By taking advantage of multilevel network structure, the hierarchical routing protocol implements a hierarchical addressing. The main idea of a routing discovery is that a routing request, which is initiated by a node at level i, is successively forwarded to its lower level $(i-1)$, until the request reaches the level that contains the destination. The hierarchical routing protocols differ in the approaches that are used for clusters to collect network state information, and the particular path used in the routing process. The hierarchical network structure reduces the number of participating nodes and the communication overhead for routing discovery and maintenance. However, these protocols suffer the communication overhead on maintaining the network hierarchy. Hierarchical state routing (HSR) [32] is a typical hierarchical routing protocol that maintains a hierarchical topology. A cluster elects a clusterhead at the lowest level and again the clusterhead is a member of the next higher level. On the higher level, superclusters are formed, and so on. A multihop route can be directly established within the cluster. On the other hand, if the destination is outside the same cluster, the node requests the clusterhead, which is able to forward the packet to the next level until to the clusterhead that the destination belongs to. The packet from this clusterhead then travels down to the destination node.

1.6.2 Multihop and Multiradio WMN Routing

The stationary MRs renders the node topology of a WMN to be fixed. On the other hand, each MR is connected to an external power supply so that it has no energy constraint. Thus, quite different from the routing design objectives in MANETs, the routing protocol in WMNs more focuses on how to determine the routing paths that can maximize the network throughput, taking advantage of multiratio. The routing protocols for MANETs have the following problems when they are directly applied in WMNs:

- *Minimum hop routing metric*. The minimum hop routing metric results in the poor throughput in WMNs. The minimum hop routing metric has been extensively used in MANETs because a route with smaller number of hops involves fewer forwarding nodes, and consequently fewer transmissions and lesser energy expense. Furthermore, a high packet delivery ratio can be guaranteed with fewer hop routes in a highly dynamical wireless environment. However, the minimum hop routing metric does not perform well in WMNs. The primary reason is that a higher packet loss or lower throughput link may be selected because of less transmission hops. The link quality is affected by two factors: the distance between the transmitter and the receiver, and the interference at the receiver. The minimum hop routing always choose the node that is furthest away as the next hop node. However, as the distance between the two nodes increases, the SNR decreases, which results in that the receiver may not be able to correctly decode the data. In addition, minimum hop metric may also select the link having the heavy interference, which also results in a low SNR in the receiver. As a result, the packet error rate and loss rate increase, which reduces the network throughput.
- *Multiradio and multichannel*. The routing protocols in MANETs are not able to take advantage of the channel diversity in WMNs. As the discussion in Sect. 1.3, multichannel and multiradio techniques improve the network capacity. However, the existing protocols in MANETs have no functionality that evaluates the impact of multiple channels or radios. In a WMN, the routing protocol has to choose the best channel or radio in terms of link quality.
- *Load-balancing route*. There is no load-balancing strategy in the MANET routing protocols. A MANET typically is dominated by peer-to-peer traffic so that the source and the destination are randomly distributed in the network. However, WMNs has to support rich Internet services and so the traffic volume may be very high and is oriented toward the IGWs. Therefore, the traffic is concentrated on the paths between the MRs and the IGWs. Furthermore, the traffic burst may occur unpredictably, depending on the applications. The MRs close the IGWs and the paths directed to the IGWs are easily turned into the hotspots because they have to deliver more Internet traffic. The minimum hop routing protocol routes the data by using the shortest path without considering load-balancing on the nodes and the paths. In this manner, when more and more packets continue arriving on a node or a path, it is overburdened, which increases the number of dropped packets and transmission delay. Moreover, additional retransmitted packets generate more traffic. As a consequence, the network performance further deteriorates.

Therefore, new routing metrics have to be designed in the WMN routing protocol that considers multiple radios and the traffic property in WMNs. In the following part, we discuss three new developed routing metrics for WMN routing.

Expected transmission count (ETX) [33] is one of the earliest new metric for WMNs. ETX takes into account asymmetry link loss in the two direction of each link and uses the fixed-size probing packets to evaluate d_{f}, the forward delivery

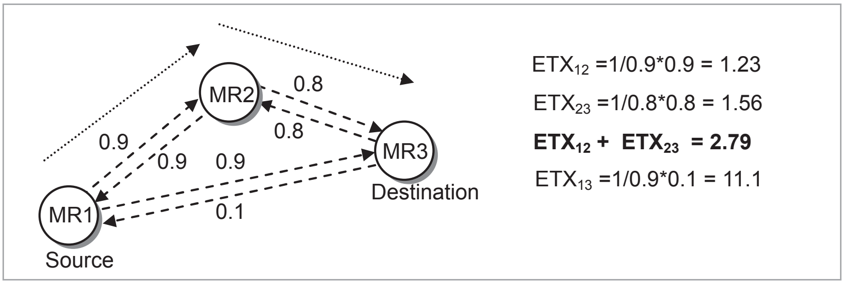


Fig. 1.7 An example of ETX

ratio, i.e., the probability that a data packet successfully arrives at the receiver, and d_r, the reverse delivery ratio, i.e., the probability that the ACK packet is successfully received. Then, the ETX value of each link is equal to $1/(d_f \times d_r)$. The EXT value of a path is the sum of the link ETX. Instead of choosing the path with minimum hops, the routing algorithm selects the path with the minimum ETX. Let us consider the source and the destination: MR1, MR3, respectively, as shown in Fig. 1.7. If a routing protocol uses the minimum hop metric as in the MANET, it chooses the direct link between MR1 and MR3 as the communication path. However, as shown in Fig. 1.7, the reverse delivery ratio from MR3 to MR1 is only 0.1, which is indicated by the successful reception probability of ACK packets. Compared to other links (i.e., 0.8), the link bandwidth in this direction is extremely low. In contrast, if the path (1–2–3) is selected using ETX as routing metric, the path can achieve a high throughput in both directions.

The drawback of EXT is that it does not identify the different throughput of the links and thus their metric may still suffer from a low throughput. On the other hand, ETX fails to consider the multichannel and traffic load balancing. These limitations motivate the development of new protocols as illustrated below.

Weighted cumulative expected transmission time (WCETT) [34] is a routing metric that takes into account both link throughput and channel diversity. It first extends the link ETX to the expected transmission time (ETT) with which to address the link bandwidth. The ETT is defined as: ETT $=$ ETX $\times$ size of probe packet/current bandwidth. Given a path with n hops, k involved channels, it defines W_i as the sum of ETT of hops working on channel i in the path:

$$W_i = \sum \mathrm{ETT}_j \quad \text{Hop } j \text{ on channel } i, \quad 1 \le i \le k. \tag{1.1}$$

Then, it defines α as the maximal value W_i of k channels

$$\alpha = \mathrm{Max}\, W_i, \quad 1 \le i \le k. \tag{1.2}$$

Furthermore, let β be the sum of ETT of all the hops of the path

$$\beta = \sum_{j=1}^{n} \mathrm{ETT}_j. \tag{1.3}$$

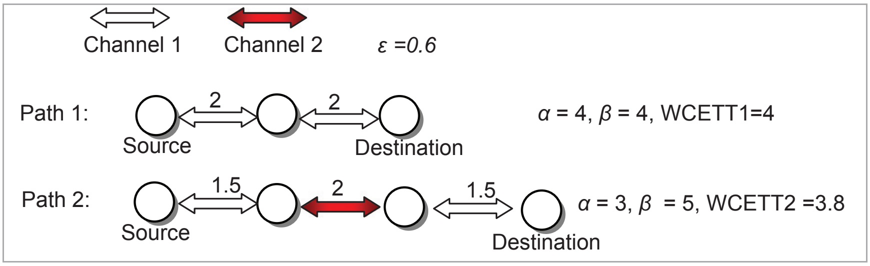


Fig. 1.8 An example of WCETT

Consequently, WCETT of a path can be formulated as

$$\text{WCETT} = (1-\varepsilon)\times\beta + \varepsilon\times\alpha, \tag{1.4}$$

where ε is a coefficient between 0 and 1.

The path with the least value of WCETT will be selected by the source node for packet transmission. The path having more channel diversity tends to have less intra interference (packets of the same flow contend with each other at different hops). Therefore, WCETT use α to capture the channel diversity. At the same time, the path with a high link quality has the lower value of β. In (1.4), ε specified the trade-off between α and β. If $\varepsilon = 0$, WCETT is similar with ETX. Figure 1.8 is an example that explains WCETT. In the example network, it has two channels, i.e., channel 1 and channel 2. Considering $\varepsilon = 0.4$, Path1 uses only one channel and has $\text{WCETT}_1 = 4$. Path 2 uses two channels. $W_{\text{channel 1}} = 1.5+1.5 = 3$, $W_{\text{channel 2}} = 2$. Thus $\alpha = 3$, $\beta = 5$ and so $\text{WCETT}_2 = 5\times 0.4 + 3\times 0.6 = 3.8$. Thus, path 2 will be selected for the source although it is a longer than path than path 1 and has the similar link quality with path 1.

However, WCETT ignores the interflow interference and the routing still cannot bypass the nodes with heavy load intentionally. How to calculate of WCETT with low expense is still an open problem.

Interference and channels switching (MIC). MIC [35] is proposed to capture interflow and intraflow interference in WMNs. The MIC of a path includes two submetrics: Interference aware resource usage (IRU) and channel switching cost (CSC). IRU evaluates the interflow interference, link quality of different bandwidth. The CSC evaluates the intraflow interference. Given a link e: it has

$$\text{IRU}_e = \text{ETT}_e \times N_e, \tag{1.5}$$

where N_e is the number of the links that can interfere e. IRU gives a larger weight to the link with more interflow interference source.

For any node i: it has

$$\text{CSC}_e = \begin{cases} w_1 & \text{No common channels} \\ w_2 & \text{Having common channels} \end{cases} \quad 0 \le w_1 < w_1. \tag{1.6}$$

$CSC_i = w_1$, if node i and its neighboring node previous node i in the path have the different channels. Otherwise, $CSC_i = w_2,\ w_1 < w_2$. It can be seen that the CSC favors the path with more channel diversity, in other words, low possible intraflow interference.

Finally, for a path p_{ij},

$$\mathrm{MIC}(p_{ij}) = \frac{1}{N \times \mathrm{Min(ETT)}} \sum_{\text{link } e \in p_{ij}} \mathrm{IRU}_e + \sum_{\text{node } x \in p_{ij}} \mathrm{CSC}_x, \tag{1.7}$$

where N is the total number of nodes in a WMN and Min(ETT) means the smallest ETT in the network.

1.7 Thoughts of Practitioners

WMNs demonstrate a strong potential of supporting high bandwidth Internet access. On the other hand, WMNs reduce the construction cost because of less dependence on the infrastructure. These critical features foster the reality civilian applications in developing cost-efficient WMNs.

The key application of WMNs is to offer Internet access at the low income community areas in a city or the rural areas to promote a variety of Internet services with a low price. An example for such kind of WMNs is TFA [7]. TFA started in 2004 to provide Internet service for the low income area of the southeast Houston. TFA covers 3 km^2 with one IGW and 18 MRs to provide 1 Mbps minimum access rate. The monthly payment for a TFA user is approximately half of the cable or DSL connections. In the early 2007, it serves 2,000 users.

Municipal high-speed Internet service is allowable by designing a WMN in the hotspot area such as mall. Users are able to access the Internet by using their laptops or PDAs. Desirable mobility can be supported by WMNs anywhere in the municipal area whether in the car or on the street. One of leading projects for such a WMN design is Chaska Net [36] being deployed in Chaska, Minnesota, USA. A number of MRs are densely mounted on the street lamps. Chaska Net plans to have 200 MRs to cover 15 square miles in the city area and maintains a low monthly charge.

Some other promising WMN applications include enterprise wireless network, metropolitan area networks, transport information system, etc.

1.8 Direction for Future Research

WMNs provide a new paradigm for high bandwidth wireless network that tightly integrates multiradio and multichannel MANET with the Internet. On the contrary to the limited civilian application of MANETs, in the past few years not only many nonprofit WMNs have been deployed, but also many industrial giants have released their commercial WMN solutions. However, there are still a number of open issues before the advantages of WMNs can fully take effect. These challenging problems involve all seven Open Systems Interconnection (ISO) protocol layers. Specifically, the critical problems related to above discussion are summarized as below:

- *Capacity improvement*. The current implemented WMNs are still far from the theoretical capacity because these implementations could not effectively combat the interference problem, channel assignment problem, etc. In order to improve the network capacity, the following issues are critically expected:
  - In the stage of network setup, the IGW and MR configuration strategy is necessary to provide the sufficient radio and channel resource for the MR, which has a high traffic demand.
  - In the stage of network deployment, the approach of distributed channel assignment is needed to dynamically assign the orthogonal channel to interfaces of MR in such a way to minimize the interference and maximize the network performance.
- *Efficient routing protocols*. Multihop, Multiradio, and Multipath routing protocols are required for effectively deploying a WMN. The key issues are:
  - WMN routing algorithm should take into account, the availability and the diversity of multiple radios and multiple channels.
  - Reliable routing metrics should be developed for efficiently identifying the link quality with the consideration of load-balancing.
- *Fairness*. Most of works pertain to fairness are still in the experimental phase and require further evaluation:
  - Enhanced or new MAC protocols should be proposed to increase the local fairness.
  - Enhanced or new transport protocol should be designed with which to effectively distinguish the packet loss because of congestion from other possible reasons in wireless networks.

1.9 Conclusions

A MANET is usually spontaneously deployed in an area for supporting peer-to-peer communication. A WMN inherits some characteristics of MANETs, such as self-origination, self-healing, and multihop communication. On the other hand, a WMN has a hierarchical architecture, i.e., static backbone and mobile clients, which facilitates the civilian application of the WMN to provide ubiquitous high-bandwidth Internet accessibility. In this chapter, we present the fundamental concepts with respect to the MANET and the WMN. In particular, we study the network design issues of a WMN from the following aspects: network capacity, IGW and MR positioning technique, unfairness transmission problem, and routing issues, by comparing them with MANETs. From these aspects we investigate the current approaches in the literature.

1.10 Terminologies

1. *Ad hoc routing*. In a MANET, each node with a limited transmission range acts as traffic source, destination, or a router to collaboratively forward data packets for other nodes without a priori knowledge of the network topology. A MANET ad hoc routing protocol is required for a node to find the communication path in a dynamic network environment. It includes two parts: routing discovery and routing maintenance.
2. *End to end unfairness*. End-to-end unfairness is that a flow that has fewer hops achieves a high throughput than a long hop flow.
3. *Internet gateway*. A few MRs with the wired connections act as the Internet gateway (IGW) to exchange the traffic between the Internet and the WMN.
4. *Local unfairness*. The phenomenon that some flows dominate the transmission for a long period time while the other flows have no chance to seize the channel for their transmission is called local unfairness.
5. *MANET*. MANET is the abbreviation of mobile ad hoc network. A pure MANET is dynamically established by mobile devices grouped together as needed without any support from the existing infrastructure. The mobile devices in the network communicate with each other through single or multihop wireless links.
6. *Mesh backbone*. The stationary wireless mesh routers (MRs) interconnecting through single/multi hop wireless links form the backbone.
7. *Network capacity*. The maximal amount of traffic load that a network can support at a time.
8. *Positioning technique*. The WMN positioning technique can be defined as physical configuration of the IGW and MRs including their locations and the number of interfaces on them.
9. *Routing metric*. Routing metric is used to differentiate the quality of different routing paths.
10. *WMN: Wireless mesh network*. A WMN is a particular multihop ad hoc network, consisting of two parts: mesh backbone, and mesh clients. The stationary wireless mesh routers (MRs) interconnecting through single/multi hop wireless links form the backbone. The MR with the wired connections acts as the IGW to exchange the traffic between the Internet and the WMN. The mesh clients can be the mobile wireless device such as cell phones and laptops. The mobile clients connect to any MRs to access the Internet via the IGW in a multihop fashion.

1.11 Questions

1. What is the main difference between MANETs and WMNs?
2. What is the main characteristic of traffic load distribution in MANETs and WMNs?
3. What is the theoretical throughput per node in a MANET of a random network model? How is the scalability of MANETs?

4. How does the network throughput improve if some infrastructure nodes are added into a MANET?
5. What are the benefits for adopting positioning technique in WMNs?
6. What is Max-flow Min-cut theorem?
7. What is the exposed terminal problem and how does the local unfairness happen in such circumstance?
8. What are the main reasons that lead to end-to-end unfairness problem?
9. What are advantages and disadvantages of reactive routing protocols in MANETs?
10. Please calculate the WCETT of the following paths. Which one will be chosen in terms of WCETT?

References

1. J. Jubin and J. D. Turnow, The DARPA packet radio network protocols, Proceedings of IEEE, 75(1), 21–32, (1987).
2. D. A. Beyer, Accomplishments of the DARPA SURAN program, Proceedings of the Military Communications Conference (MILCOM), Sep (1990).
3. C. E. Perkins and E. M. Royer, Ad-hoc on-demand distance vector routing, Proceeding of the Second IEEE workshop Mobile Computing System and Applications, 90–100 Feb (1999).
4. D. B. Johnson and D. A. Maltz, Dynamic Source Routing in Ad-Hoc Wireless Networks, Mobile Computing, T. Imielinski and H. Korth, Eds., Kluwer, Dordrecht, 153–181 (1996).
5. J. Bicket, D. Aguayo, S. Biswas, and R. Morris, Architecture and evaluation of an unplanned 802.11b mesh network, Proceedings of the Eleventh Annual International Conference on Mobile Computing and Networking (Mobicom), Aug (2005).
6. S. A. Mahmud, S. Khan, S. Khan, and H. Al-Raweshidy, A comparison of MANETs and WMNs: Commercial feasibility of community wireless networks and MANETs, Proceeding of the First International Conference on Access Networks (AccessNet), Athens, Greece, Sep (2006).
7. J. Camp, E. Knightly, and W. Reed, Developing and deploying multihop wireless networks for low-income communities, Proceedings of Digital Communities, Jun (2005).
8. http://www.dailywireless.org/2007/04/23/belair-live-in-london/
9. http://www.cisco.com/en/US/products/ps6548/prod_brochure0900aecd8036884a.html
10. P. Gupta and P. R. Kumar, The capacity of wireless networks, IEEE Transactions on Information Theory, 46(2), 388–404, (2000).
11. B. Liu, Z. Liu, and D. Towsley, Capacity of a wireless ad hoc network with infrastructure, Computer Science Dept. University of Massachusetts Amherst, Technical Report, (2004).
12. P. Kyasanur and N. H. Vaidya, Capacity of multi-channel wireless networks: Impact of number of channels and interfaces, Proceedings of the Eleventh Annual International Conference on Mobile Computing and Networking (Mobicom), Aug (2005).
13. Y. Bejerano, Efficient integration of multihop wireless and wired networks with QoS constraints, IEEE/ACM Transaction on Networking, 12(6), 1064–1078, (2004).
14. L. Qiu, R. Chandra, K. Jain, and M. Mahdian, Optimizing the placement of integration points in multi-hop wireless networks, Proceeding of the Twelfth IEEE International Conference on Network Protocols (ICNP), Oct (2004).
15. R. Prasad and H. Wu, Minimum-cost gateway deployment in cellular WiFi networks, Proceeding of Consumer Communications and Networking Conference(CCNC), Las Vegas, Jan (2006).
16. B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, Gateway placement optimization in wireless mesh networks with QoS constraints, IEEE Journal on Selected Areas in Communications, 24(11), 2127–2136, (2006).

17. B. He, B. Xie, and D. P. Agrawal, Optimizing the internet gateway deployment in a wireless mesh network, Proceeding of the Fourth IEEE International Conference on Mobile Ad-Hoc and Sensor Systems(MASS), Pisa, Italy, Oct (2007).
18. A. Raniwala and T-c. Chiueh, Architecture and algorithms for an IEEE 802.11-based multi-channel wireless WMN, IEEE INFOCOM, Mar (2005).
19. J. Wang, B. Xie, K. Cai, and D. P. Agrawal, Efficient mesh router placement in wireless mesh networks, Proceeding of the Fourth IEEE International Conference on Mobile Ad-Hoc and Sensor Systems(MASS), Pisa, Italy, Oct (2007).
20. K. Cai, M. Blackstock, R. Lotun, M. Feeley, C. Krasic, and J. Wang, Wireless unfairness: Alleviate MAC congestion first!, Proceeding of the Second ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization (WiNTECH) in Conjunction with MobiCom, Montreal, Canada, Sep (2007).
21. T. Nandagopal, S. Lu, and V. Bharghavan, A unified architecture for the design and evaluation of wireless fair queueing algorithms, Proceedings of the Fourth Annual International Conference on Mobile Computing and Networking (Mobicom), Oct (1998).
22. N. H. Vaidya, P. Bahl, and S. Gupta, Distributed fair scheduling in a wireless LAN, Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking (Mobicom), Aug (2000).
23. F. Talucci, M. Gerla, and L. Fratta, Macabi (maca by invitation): A receiver oriented access protocol for wireless multiple networks, in PIMRC 1997, 1994, pp. 1–4.
24. B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, Opportunistic media success for multirate ad hoc networks, Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking (Mobicom), Sep (2002).
25. V. Gambiroza, B. Sadeghi, and E. Knightly; End-to-end performance and fairness in multihop wireless backhaul networks, Proceedings of the Tenth Annual International Conference on Mobile Computing and Networking (Mobicom), Sep (2004).
26. A. Raniwala, P. De, S. Sharma, R. Krishnan, and T. Chiueh, End-to-End flow fairness over IEEE 802.11-based wireless mesh networks, INFOCOM, May (2007).
27. K. Sundaresan, V. Anantharaman, H. Y. Hsieh, and R. Sivakumar; ATP: A reliable transport protocol for ad hoc networks, Proceeding of the Fourth ACM Interational Symposium on Mobile Ad Hoc Networking and Computing(MobiHoc), Jun (2003).
28. K. Chen, K. Nahrstedt, and N. Vaidya; The utility of explicit rate-based Flow control in mobile ad hoc networks, Proceeding of IEEE Wireless Communications and Networking Conference (WCNC), Mar (2004).
29. C. E. Perkins and P. Bhagwat, Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers, Computer Communication Review, 24(4), 234–244, (1994).
30. S. Murthy and J. J. Garcia-Luna-Aceves, An efficient routing protocol for wireless networks, ACM Mobile Networks and Appilications Journal, Special Issue on Routing in Mobile Communication Networks, 1, 183–197, (1996).
31. D. P. Agrawal and Qing-An Zeng, Introduction to Wireless and Mobile Systems, Chapter 13, Brooks/Cole (Thomson Learning), Pacific Grove, CA, ISBN No. 0534-40851-6
32. A. Iwata, C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, Scalable routing strategies for ad hoc wireless networks, IEEE Journal on Selected Areas in Communications, Special Issue on Ad-Hoc Networks, 17(8), 1369–1379, (1999).
33. D. De. Couto, D. Aguayo, J. Bicket, and R. Morris, High-throughput path metric for multi-hop wireless routing, Proceedings of the Ninth Annual International Conference on Mobile Computing and Networking (Mobicom), Sep (2003).
34. R. Draves, J. Padhye, and B. Zill, Routing in multi-radio, multi-hop. wireless mesh networks, Proceedings of the Tenth Annual International Conference on Mobile Computing and Networking (Mobicom), Sep (2004).
35. Y. Yang, J. Wang, and R. Kravets, Designing routing metrics for mesh networks, Proceeding of IEEE Workshop on Wireless Mesh Networks (WiMesh), Sep (2005).
36. http://www.chaska.net/
37. B. M. Leiner, R. J. Ruth, and A. R. Sastry, Goals and challenges of the DARPA GloMo program [global mobile information systems], IEEE Personal Communications, 3(6), 34–43, (1996).