CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

1 Introduction

The explosive growth and widespread applications of high-rate multimedia wireless communications have prompted waves of research and standard development activities. The ultimate goal is to connect all users to their targets all the time: to the Internet, to the cloud, and to the various devices in their lives, from phones to PC to tablet and to their cars. They expect to access a rich set of services regardless of where they are: whether at home, around the office, or outdoors. This exponential growth in data traffic is inevitably limited by spectrum availability and energy consumption of mobile terminals and access networks. The future success of wireless networks thus relies on the ability to overcome the mismatch between the requested *quality-of-service* (QoS) and limited network resources.

1.1 Motivation

Among available resources, spectrum and energy are two fundamental ones that enable wireless communications. Spectrum is a natural resource that cannot be replenished and therefore must be used efficiently; so that is where the special significance of *spectral efficiency* lies. Tremendous efforts and progress have been made in past decades to improve spectrum efficiency of wireless networks. However, the increasing market for all things wireless generates more demand than supply can handle. Wireless networks therefore need to be designed in a much more spectrum-efficient way to keep up with consumer demands.

Indeed, *energy efficiency* is also becoming increasingly important. From the perspective of user experiences, small form factor mobile devices are getting more and more energy hungry, since battery technology has not kept up with the growing requirements stemming from ubiquitous multi-media applications [114]. From a global perspective, we are confronted with severe challenges of environmental protection and prevention of climate changes. Improving the energy efficiency of wireless networks will not only reduce their impacts on environments but also cut network costs and make communications more affordable for everyone [68].

Spectral efficiency and energy efficiency are two different metrics for measuring wireless network efficiency. Some design criteria optimized for improving one metric may not necessarily improve the other. For example, a single wireless user may achieve

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

its highest spectral efficiency if it transmits with the highest radio power. However, energy efficiency will then be adversely affected due to too high or concentrated consumption of that energy resource. We can see that there is an urgent need to address spectral efficiency and energy efficiency issues together. New metrics, which are completely distinct from existing ones in literature, may be required to address this need. The spectral efficiency–energy efficiency relationship varies depending on the type of wireless networks we design. Different approaches should therefore be analyzed and better understood in order to design spectral- and energy-efficient operations for different types of wireless networks. This will be the focus of this book.

1.2 Wireless networks

1.2.1 Overview

Wireless networks refer to communication networks of any kind that are implemented by using radio communications. Wireless networks have various application requirements, including coverage, data rate, delay, error rate, mobility, functionality, and so on. No single wireless technology is capable of handling all these requirements effectively, and that is why different types of wireless networks with different topologies and coverage have been designed to handle different situations. The protocols running on different types of wireless networks also differ significantly from one another.

In general, wireless networks can be classified into two main categories, having or not having infrastructure support. Wireless networks with infrastructure support typically consist of fixed access points, e.g. base stations, and wireless mobile users (or nodes, devices, mobile stations, which we will use interchangeably). The access points are connected to the Internet via backhaul links and the mobile users communicate with the access points through wireless air interfaces. Some typical infrastructure wireless networks are cellular networks and wireless local area networks (WLANs). For those without infrastructure support, e.g. a mobile ad hoc network, the network can operate autonomously. In a mobile ad hoc network, each user can communicate with neighbors within communication range, and may operate as a source, destination, router, or relay. Because they are easy to deploy, ad hoc networks can be applied in critical situations such as battlefield communications and disaster recovery, where infrastructure networks are hard to build or maintain.

Traditionally, wireless networks have limited or no inter-operability between the various technologies. With the advances in wireless technologies, the evolution is toward a mixture of various technologies and topologies that coexist and inter-operate to provide seamless services to mobile users anywhere. Future wireless networks will consist of hybrid structures and different access technologies. For example, they may consist of both wide-area cellular networks providing high-speed mobile services to users over a large area, small cells covering local area hot spots with much higher data rates, and ad hoc direct communications among users in a vicinity. Correspondingly, mobile users in such a hybrid network are expected to operate with multiple access protocols and switch intelligently in different operating modes.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

As multiple users need to share limited resources, resource management essentially guarantees the performance of individual users as well as that of the whole network. There are well-known orthogonal wireless resources in four dimensions; time, frequency, space, and code. As long as different transmissions occur in resources that have no overlap in any of the four dimensions, they do not interfere with each other. So the problem of resource management is how many orthogonal resource chunks should be assigned to each user. Another resource, that is even more fundamental, is energy, which is a non-orthogonal resource and is much more difficult to manage. When different users send data over different time, frequency, space, or code dimensions, their power can be managed independently. However, when the transmissions overlap in time, frequency, space, or code, the transmissions interact with each other through mutual interference. For example, inter-cell co-channel interference in cellular networks happens because there is overlap in the space domain and different cells are not spatially separated far enough. Inter-symbol interference takes place when there is time overlap and previous symbols are delayed and overlap with the following symbols. Similarly, inter-channel interference in frequency division multiple access (FDMA) and code of division multiple access (CDMA) result from overlap in frequency or code. Interference is controlled by transmission power levels. In addition, transmission power also determines energy consumption. Therefore, power plays a pivotal role in both network spectrum and energy efficiency.

There are many other factors affecting wireless resource management. One principal limitation is the wireless channel. With wireless communications, transmission signals are usually severely distorted by the channel and the distortions differ from user to user and from time to time. The wireless channels of different users also vary significantly in the number of paths, path delays, phase shifts, path attenuation, etc. Furthermore, the broadcast nature of wireless channels implies limitations on transmission power to avoid co-channel interference, because excessive interference can deteriorate network performance and waste scarce wireless resources. In addition, government regulations also apply strict control on the amount of power that can be transmitted, such that the radio frequency (RF) exposure levels people may be subjected to are safe. As a consequence, limited power can be used for wireless transmission to compensate the path loss. The wireless channel is therefore error-prone and highly unreliable and subject to many impairment factors that are of transient nature.

Besides these physical limitations, users also have different QoS demands, indicating the need to use different amounts of wireless resources. The interest of one user usually conflicts with that of others in resource allocation. Therefore, it is critical to design resource management schemes that can allocate wireless resources effectively, efficiently, and fairly. It is the scheduling and medium access control (MAC) protocols that allocate wireless resources to users on demand, multiplex and separate transmissions of different users, control interference, and ensure network-wide flexibility, efficiency, and fairness of resource sharing. These schemes need to consider both wireless properties and user demands. The diversity underlying channel conditions and user demands should be exploited together to enhance network performance.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

Wireless resource management schemes can be divided into several categories depending on if they use central controllers or not. In centralized networks, a central scheduler collects all network information and decides the resource allocation for all users in the network. In distributed wireless networks, users in the network make decisions on resource allocation themselves and the conflicts in resource allocation among users are resolved by their autonomous behaviors. The performance of distributed approaches can be enhanced through distributed collaborations among users in the network. The centralized and distributed schemes can be combined together to further improve the spectral and energy efficiency of the network.

1.2.2 Traditional layered architecture

Traditionally, network protocols are described with a layered model and the protocols are stacked on top of each other. Each layer has specific responsibilities and knows nothing about the procedures of other layers. Each layer carries out its own tasks and delivers messages to the adjacent layer in the process. Data sent to others are passed from the highest-level protocol down to the lowest-level one and vice versa. The layered model allows network services to be defined with their functions, rather than specific implementations. The isolation of communications functions in different layers minimizes the impact of technique change on the entire protocol stack. Protocols in each layer can be updated without affecting other layers, as long as the updates use the same interfaces with adjacent layers as those used before.

A classic layered network model developed by the International Standards Organisation (ISO) is called the Open Systems Interconnect (OSI) reference model, which is frequently used to describe the structure and function of data communication protocols. The OSI model contains seven layers and each layer represents a function performed when data are transferred through the layer. In the OSI model, a layer defines the function instead of the protocols used to implement the function. Therefore, each layer may have multiple protocols. Every layer communicates only with its remote peer entity, which runs the same protocol in the equivalent layer.

The OSI model has seven layers, which are described briefly in the following:

- Layer 1: Physical layer. This layer defines the functions of the hardware necessary to transmit the data signals on a certain carrier. The main responsibilities of this layer are to send and receive information bits to or from the medium. It describes the way data are actually transmitted on the channel, but does not define the medium. There are many varieties of media for data communication, e.g. radio, cable, fibre optics, light waves, and so on. Different medium need different sets of physical layer protocols. The physical layer describes how information bits are encoded into media signals and the characteristics of the media interface.
- Layer 2: Link layer. This layer transfers data between entities in the network, detects, and possibly corrects errors that may occur in the physical layer. It formats packets, defines network-frames, and is responsible for error control on the frame level. Not all physical layer bits go into link layer frames, as

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

some of these bits are for physical layer functions. The link layer connects users in the same network and provides intranet address information for the physical layer in the transmitted frames. Usually the link layer is divided into two sublayers: the MAC and logical link control (LLC) layers. The MAC sublayer controls how users in the network gain access to the channel to transmit data, while the LLC layer controls frame synchronization, flow control, and error checking.

- Layer 3: Network layer. This layer transmits data and decides which route the data must follow through the whole network. The network layer maintains the quality of service provided to the upper transport layer. The network layer receives data packets from the upper layer from the sender, and transmits them by as many connections and subsystems as needed to reach the destination. The source node may reside in one type of network while the destination in a different one. This layer also controls packet delivery between intermediate stations. Therefore, the network layer manages connections across the network and isolates the upper-layer protocols from the details of the underlying physical networks. An example is the Internet Protocol (IP), which handles the addressing and delivery of data in the TCP/IP protocol.
- Layer 4: Transport layer. This layer ensures transparent transfer of data between end users and is responsible for end-to-end error recovery and flow control. It controls the reliability through flow control, segmentation/desegmentation, and error control. The transport layer may keep track of the segments of a packet and retransmit those that fail to ensure reliable delivery. Sometimes the transport layer also provides acknowledgement of successful data transmission. In TCP/IP, the function of the transport layer is performed by TCP. However, TCP/IP offers a second transport layer protocol, the User Datagram Protocol (UDP) that does not perform the end-to-end reliability checks.
- Layer 5: Session layer. This layer establishes, manages, and terminates connections between dialogues or connections of users. A dialogue is a formal conversation in which two nodes agree to exchange data. It deals with session and connection coordination. The communication can take place in full-duplex, half-duplex, or simplex mode. Sessions enable users to communicate in an organized manner. Each session has three phases: connection establishment, data transfer, and connection release. In the establishment phase, the source and destination users negotiate the rules of communication, including the protocols to be used and communication parameters. Then they exchange data in a dialogue. When the users no longer need to communicate, they engage in an orderly release of the session.
- Layer 6: Presentation layer. This layer cooperates applications to exchange data and is responsible for presenting data to the application layer. It transforms data into the form that the application layer can accept and provides freedom from compatibility problems. For example, translations could be made between ASCII and Unicode. In addition, this layer may also provide security assurance in the form of encryption and compression.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

- Layer 7: Application layer. This layer supports applications that users directly interact with, as well as other processes that users are not necessarily aware of. For example, this layer provides application services for file transfers, e-mail, and other network services. Telnet, FTP, and Email are typical examples.

One advantage of the layered protocol is that they break the communication process into manageable chunks. Designing a small part of the protocol stack is much easier than designing the entire model. So it simplifies engineering. A change in one layer does not affect others and new technology can be easily introduced into the system.

1.2.3 Necessity of cross-layer optimization

As introduced in the previous section, communication networks can be modeled using the OSI standard and the networks are divided into layers that are in charge of different functionalities. For example, the physical (PHY) layer takes care of reliable and efficient bit transmission using modulation and coding techniques. The MAC layer handles resource allocation to multiple users. On the other hand, the network layer is in charge of routing. The traditional design paradigm emphasizes transparency between layers for the purpose of implementation simplicity. With this design, each layer does not need to know how adjacent layers work inside. Instead, each layer accesses only the interfaces provided by the adjacent layers and the interfaces are usually minimally designed. The layered design has some advantages from a design and implementation perspective. For example, each layer can be independently updated without affecting other layers. However, the layered design ignores coupling among adjacent layers that leads to significant information loss between layers and thus significant performance degradation. The performance loss is even more significant in wireless networks than wired ones. This is because the MAC layer is closely related to the underlying physical layer. For example, in the downlink of a wireless cellular network, the network sum capacity is maximized if, in each time slot, the user with the best instantaneous channel gain is scheduled. This is one way of exploiting the so-called multi-user diversity, and the scheduler is a channel-aware scheduler. The instantaneous channel gain is PHY layer knowledge. However, the MAC layer deals with user scheduling. With traditional layered protocol, the MAC layer will not be able to apply the channel-aware scheduler to exploit multi-user diversity because of the lack of channel information. If there are many users in the network and the channel varies significantly among users, the loss in spectral and energy efficiency would be huge without exploiting multi-user diversity.

Spectral and energy efficiency are affected by all layers of system design, ranging from silicon to applications. While the traditional layer-wise approach leads to independent design of layers and results in high design margin, cross-layer approaches exploit interactions between different layers and can significantly improve system performance as well as adaptability to service, traffic, and environment dynamics. Cross-layer optimization for throughput improvement has been a popular research theme [197, 151, 131]. Recent efforts have also been undertaken to tackle energy consumption

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

at all layers of communication systems, from architectures [135, 119, 29] to algorithms [48, 93, 216].

The PHY layer plays a very important role in wireless communications due to the challenging nature of the communication medium. In wireless networks, the PHY layer deals with data transmission over wireless channels and consists of RF circuits, modulator, power control, channel coding units, etc. The physical layer has many tasks and some examples are listed below.

- Modulation and demodulation: Map information bits into analog signals and vice versa. Some common techniques are:
  - Amplitude modulation (AM): Uses different amplitude levels of the carrier signal to represent information bits. AM is not robust to noise and interference and rarely used in wireless communications.
  - Frequency modulation (FM): Uses different frequencies to represent information bits.
  - Phase shift keying (PSK): Uses the phase of the carrier signal to represent information bits.
- Coding and decoding: Convert messages from their original forms, e.g. bits, into other forms that represent the messages for efficient transmission and vice versa. There are numerous encoding and decoding algorithms and the oldest one is the Morse code, which was used in the landline telegraph in the 19th century.
- Time or frequency synchronization: Enable the transmitter and receiver to agree on the frequency or time that the communications take place.
- Multiplexing and demultiplexing: Allow multiple users to transmit at the same time without interfering with each other. For example, with frequency division multiplexing (FDM), different users will use different sets of frequencies to send data at the same time.
- Carrier sensing in some MAC protocols: Detects if the carrier is in use before attempting to send data.
- Signal processing: Uses equalization, filtering, pulse shaping, channel estimation, signal detection, and so on, that are used to process signals.
- Interleaving: Reorders data such that consecutive bits are distributed over a larger sequence of data to reduce burst errors. Many error protection coding algorithms cannot correct for errors in groups, and interleaving increases their ability to correct for burst errors.

Traditional wireless systems are built to operate on a fixed set of operating points [38], e.g. no power adaptation. This results in excessive energy consumption or a pessimistic data rate for peak channel conditions. Hence, a set of PHY parameters should be adjusted to adapt the actual user requirements (e.g. throughput and delay) and environments (such as shadowing and frequency selectivity) to tradeoff energy efficiency and spectral efficiency. As wireless is a shared medium, communication performance and energy consumption are affected not only by the layers comprising the point-to-point communication link, but also by the interaction between the links in the entire network. Hence, a system approach is required.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

On the other hand, in a multi-user network, the MAC layer ensures that wireless resources are efficiently allocated to maximize network-wide performance metrics while maintaining user QoS requirements. Here, pessimistic medium access strategies that allocate wireless resources to assure worst-case QoS may hurt network spectral and energy efficiency. In distributed access schemes, MAC should be improved to reduce the number of wasted transmissions that are corrupted by interference from other users, while in centralized access schemes, efficient scheduling algorithms should exploit the variations across users to maximize the overall network performance. The MAC layer manages wireless resources for the PHY layer and they both directly impact overall network performance and energy consumption.

In this book, we introduce cross-layer approaches to optimize the performance of wireless networks. More specifically, we emphasize joint physical-MAC layer designs to improve wireless spectral and energy efficiency because the two layers closely depend on each other. We will emphasize scheduling, channel access, radio power control, modulation, coding, network energy consumption, and their interactions, together with the wireless channel states.

Orthogonal frequency division multiplexing (OFDM) is a key modulation scheme for next-generation broadband wireless standards [63, 18], including digital video broadcasting (DVB) systems, WLAN standards such as American IEEE 802.11 a/g/n and the European equivalent HIPERLAN/2, the fourth-generation (4G) mobile communications such as IEEE WiMAX and 3GPP LTE. OFDM has been widely applied in wireless networks because of its high bandwidth efficiency and robustness to multi-path fading and delay channels. In addition, OFDM converts a frequency-selective fading channel into several nearly flat-fading channels by dividing the entire available spectrum into narrow-band subchannels. The high spectral efficiency is obtained by overlapping the orthogonal frequency responses of the subchannels. From a resource allocation perspective, these multiple channels in OFDM systems have the potential for more efficient MAC design since subcarriers can be assigned to different users [8] and this is usually called Orthogonal Frequency-Division Multiple Access (OFDMA). Furthermore, adaptive power allocation on each subcarrier can be applied for further improvement [9]. Therefore, the exploitation of these OFDMA properties in network resource management will significantly boost network spectral and energy efficiency. This book will emphasize joint PHY and MAC optimization to improve the spectral and energy efficiency for OFDM-based wireless networks.

1.3 Book outline

The major goal of this book is to introduce state-of-the-art cross-layer transmission and resource management designs that significantly improve both spectral efficiency and energy efficiency in wireless networks. The book is divided into four parts. In Part I, we introduce the basic concepts of wireless communications and networks that serve as a foundation for understanding the book. Readers familiar with this background knowledge can skip this part and start from Part II directly. Part II introduces cross-layer

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

designs for networks with central controllers while Part III does so for networks without central controllers. Both Parts II and III emphasize enhancing spectral efficiency. In Part IV, energy-efficient design techniques are introduced and the relations between energy efficiency, spectral efficiency, and some other network performance metrics as well as their impact on network designs are discussed in detail.

To be more specific, Part I consists of four chapters. The first chapter talks about wireless channel properties such as path loss, shadowing, and fading for individual links. Channel state information (CSI) is essential in wireless networks for cross-layer design and we will discuss some channel estimation methods that can be used to obtain the CSI. The second chapter gives a detailed comparison of the spectral and energy efficiency metrics from both link and network levels. In the third and fourth chapters, we discuss traditional MAC protocols for different types of wireless networks. We will introduce both centralized resource management schemes and distributed access protocols.

Part II is focused on centralized cross-layer optimization assuming a centralized scheduler to manage network resources. It uses two major mechanisms in resource management to improve the spectral efficiency of wireless networks: exploiting the time variance and frequency selectivity of wireless channels through adaptive modulation, coding, as well as packet scheduling and regulating resource allocation through network economics. With the help of utility functions that capture the satisfaction level of users for a given resource assignment, Part II introduces a generic utility optimization framework for centralized OFDM networks, in which the network utility at the level of applications is maximized subject to the current channel conditions and the modulation and coding techniques employed in the network. We will introduce novel efficient dynamic subcarrier assignment (DSA) and adaptive power allocation (APA) algorithms, which are proven to achieve optimal or near-optimal performance with very low complexity. Based on a holistic design principle, we will also introduce max-delay-utility (MDU) scheduling, which senses both channel and queue information. MDU scheduling can simultaneously improve spectral efficiency and provide the right incentives to ensure that all applications can receive the different required QoS. To facilitate cross-layer design, we will also investigate the mechanisms of channel-aware scheduling, such as efficiency, fairness, and stability, using extreme value theory. Especially, we will analyze the impact of multi-user diversity on throughput and packet delay and provide a method to design cross-layer scheduling algorithms that allow the queueing stability region at the network layer to approach the ergodic capacity region at the physical layer.

In Part III, we continue the discussion of spectrum efficiency improvement using cross-layer techniques but in distributed wireless networks. We first introduce the concept of distributed multi-user diversity and its potential to improve network throughput and then the detailed technologies to exploit this diversity. The design philosophy of distributed approaches heavily depends on how different users in the network interact or interfere with others. When different communicating pairs are very close to each other, they will not be able to send data simultaneously and distributed MAC protocols are needed to avoid collisions. We will first introduce distributed MAC protocols that utilize channel state information to exploit the multi-user diversity in the wireless channels for collision resolutions. In these protocols, we will discuss opportunistic random access

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-03988-9 - Energy and Spectrum Efficient Wireless Network Design
Guowang Miao and Guocong Song
Excerpt
More information

for single-cell cellular networks and then for any network topologies. The final goal is to introduce how to use distributed random access approaches to achieve performance comparable to that of centralized approaches when channel state information is used at the MAC layer. We will introduce an optimal channel-aware distributed MAC, which, although quite preliminary, reaches this goal. In the last chapter, as an example, we will discuss how in practice we can use these distributed approaches in cellular networks to improve spectral efficiency. Besides resolving collisions, power control is essential in determining network performance for simultaneous data transmissions. At the end of this part, we will introduce distributed power control for both real-time and elastic traffic.

In Part IV, the focus is on energy-efficient cross-layer design techniques, and the relationship between energy and spectrum efficiency in various types of wireless networks. We will use a bottom-up approach. This means we will first study simple point-to-point energy-efficient communications and then more complicated multi-user networks. We start by discussing energy-efficient wireless transmission techniques in both flat-fading and frequency-selective channels. After that, we consider a multi-user single-cell network and discuss energy-efficient orthogonal resource management schemes in different resource domains. We will introduce energy-efficient scheduling technologies both with and without fairness and show that conventional spectrum-efficient schedulers are indeed special cases of energy-efficient schedulers in specific regimes. Then we will study distributed energy-efficient communications, including both energy-efficient distributed random access and power control in interference-limited networks. We will discuss the fundamental tradeoffs in wireless resource allocation and investigate relationships between energy efficiency, spectral efficiency, and several other network performance metrics such as deployment cost, system bandwidth, etc. Finally, we move on to the whole network level and investigate system-level energy-efficient designs for both homogeneous and heterogeneous cellular networks. We will discuss implementation issues in practice, illustrating how the technologies discussed in this part can be implemented in real-world wireless networks.

© in this web service Cambridge University Press www.cambridge.org