MIT 6.02 **DRAFT** *Lecture Notes*
Last update: September 6, 2011
Comments, questions or bug reports?
Please contact hari at mit.edu

CHAPTER 1
Introduction

Our mission is to expose you to a variety of different technologies and techniques in electrical engineering and computer science. We will do this by studying several salient properties of **digital communication systems**, learning both important aspects of their design, and also the basics of how to analyze their performance. Digital communication systems are well-suited for our goals because they incorporate ideas from a large subset of electrical engineering and computer science.

Equally important, the ability to disseminate and exchange information over the world's communication networks has revolutionized the way in which people work, play, and live. At the turn of the century when everyone was feeling centennial and reflective, the U.S. National Academy of Engineering produced a list of 20 technologies that made the most impact on society in the 20th century.[1] This list included life-changing innovations such as electrification, the automobile, and the airplane; joining them were four technological achievements in the area of commununication—*radio and television*, the *telephone*, the *Internet*, and *computers*—whose technological underpinnings we will be most concerned with in this book.

Somewhat surprisingly, the Internet came in only at #13, but the reason given by the committee was that it was developed toward the latter part of the century and that they believed the most dramatic and significant impacts of the Internet would occur in the 21st century. Looking at the first decade of this century, that sentiment sounds right—the ubiquitous spread of wireless networks and mobile devices, the advent of social networks, and the ability to communicate any time and from anywhere are not just changing the face of commerce and our ability to keep in touch with friends, but are instrumental in massive societal and political changes.

Communication is fundamental to our modern existence. Who among you can imagine life without the Internet and its applications and without some form of networked mobile device? Most people feel the same way—in early 2011, over 5 billion mobile phones were active worldwide, over a billion of which had "broadband" network connectivity. To put this number (5 billion) in perspective, it is larger than the number of people in the world

---

[1] "The Vertiginous March of Technology", obtained from nae.edu. Document at `http://bit.ly/owMoO6`

who (in 2011) have electricity, shoes, toothbrushes, or toilets![2]

What makes our communication networks work? This course is a start at understanding the answers to this question. This question is worth studying not just because of the impact that communication systems have had on the world, but also because the technical areas cover so many different fields in EECS. Before we dive in and describe the "roadmap" for the course, we want to share a bit of the philosophy behind the material.

Traditionally, in both education and in research, much of "low-level communication" has been considered an "EE" topic, covering primarily the issues governing how bits of information move across a single communication link. In a similar vein, much of "networking" has been considered a "CS" topic, covering primarily the issues of how to build communication networks composed of multiple links. In particular, many traditional courses on "digital communication" rarely concern themselves with how networks are built and how they work, while most courses on "computer networks" treat the intricacies of communication over physical links as a black box. As a result, a sizable number of people have a deep understanding of one or the other topic, but few people are expert in every aspect of the problem. As an abstraction, however, this division is one way of conquering the immense complexity of the topic, but our goal in this course is to both understand the important details, and also understand how various abstractions allow different parts of the system to be designed and modified without paying close attention (or even really understanding) what goes on elsewhere in the system.

One drawback of preserving strong boundaries between different components of a communication system is that the details of how things work in another component may remain a mystery, even to practising engineers. In the context of communication systems, this mystery usually manifests itself as things that are "above my layer" or "below my layer". And so although we will appreciate the benefits of abstraction boundaries in this course, an important goal for us is to study the most important principles and ideas that go into the complete design of a communication system. Our goal is to convey to you both the breadth of the field as well as its depth.

In short, we cover communication systems all the way from the *source*, which has some information it wishes to transmit, to *packets*, which messages are broken into for transmission over a network, to *bits*, each of which is a "0" or a "1", to *signals*, which are analog waveforms sent over physical communication links (such as wires, fiber-optic cables, radio, or acoustic waves). We describe the salient aspects of all the layers, starting from how an application might encode messages to how the network handles packets to how links manipulate bits to how bits are converted to signals for transmission. In the process, we will study networks of different sizes, ranging from the simplest *dedicated point-to-point link*, to *shared media* with a set of communicating nodes sharing a common physical communication medium, to larger *multi-hop networks* that themselves are connected to other networks to form even bigger networks.

---

[2] It is in fact distressing that according to a recent survey conducted by TeleNav—and we can't tell if this is a joke—40% of iPhone users say they'd rather give up their toothbrushes for a week than their iPhones! `http://www.telenav.com/about/pr-summer-travel/report-20110803.html`

1.1 Themes

Three fundamental challenges lie at the heart of all digital communication systems and networks: *reliability*, *sharing*, and *scalability*. We will spend a considerable amount of time on the first two issues in this introductory course, but much less time on the third.

1.1.1 Reliability

A large number of factors conspire to make communication unreliable, and we will study numerous techniques to improve reliability. A common theme across these different techniques is that they all use redundancy in creative and efficient ways *to provide reliability using unreliable individual components*, using the property of independent (or perhaps weakly dependent) failures of these unreliable components to achieve reliability.

The primary challenge is to overcome a wide range of faults and disturbances that one encounters in practice, including *Gaussian noise* and *interference* that distort or corrupt signals, leading to possible *bit errors* that corrupt bits on a link, to *packet losses* caused by uncorrectable bit errors, *queue overflows*, or *link and software failures* in the network. All these problems degrade communication quality.

In practice, we are interested not only in reliability, but also in speed. Most techniques to improve communication reliability involve some form of redundancy, which reduces the speed of communication. The essence of many communication systems is how reliability and speed tradeoff against one another.

Communication speeds have increased rapidly with time. In the early 1980s, people would connect to the Internet over telephone links at speeds of barely a few kilobits per second, while today 100 Megabits per second over wireless links on laptops and 1-10 Gigabits per second with wired links are commonplace.

We will develop good tools to understand why communication is unreliable and how to overcome the problems that arise. The techniques involve error-correcting codes, handling distortions caused by "inter-symbol interference" using a *linear time-invariant* channel model, *retransmission protocols* to recover from packet losses that occur for various reasons, and developing *fault-tolerant routing protocols* to find alternate paths in networks to overcome link or node failures.

1.1.2 Efficient Sharing

"An engineer can do for a dime what any fool can do for a dollar," according to folklore. A communication network in which every pair of nodes is connected with a dedicated link would be impossibly expensive to build for even moderately sized networks. *Sharing* is therefore inevitable in communication networks because the resources used to communicate aren't cheap. We will study how to share a point-to-point link, a shared medium, and an entire multi-hop network among multiple communications.

We will develop methods to share a common communication medium among nodes, a problem common to wired media such as broadcast Ethernet, wireless technologies such as wireless local-area networks (e.g., 802.11 or WiFi), cellular data networks (e.g., "3G"), and satellite networks (see Figure 1-1).

We will study modulation and demodulation, which allow us to transmit signals over different carrier frequencies. In the process, we can ensure that multiple conversations

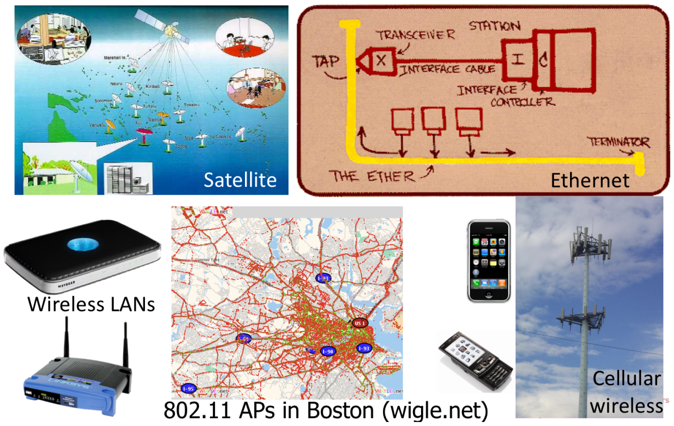

Figure 1-1: Examples of shared media.

share a communication medium by operating at different frequencies.

We will study *medium access control* (MAC) protocols, which are rules that determine how nodes must behave and react in the network—emulate either *time sharing* or *frequency sharing*. In time sharing, each node gets some duration of time to transmit data, with no other node being active. In frequency sharing, we divide the communication bandwidth (i.e., frequency range) amongst the nodes in a way that ensures a dedicated frequency sub-range for different communications, and the different communications can then occur concurrently without interference. Each scheme has its sweet spot and uses.

We will then turn to *multi-hop* networks. In these networks, multiple concurrent communications between disparate nodes occurs by sharing over the same links. That is, one might have communication between many different entities all happen over the same physical links. This sharing is orchestrated by special computers called *switches*, which implement certain operations and protocols. Multi-hop networks are generally controlled in distributed fashion, without any centralized control that determines what each node does. The questions we will address include:

1. How do multiple communications between different nodes share the network?
2. How do messages go from one place to another in the network—this task is faciliated by *routing* protocols.
3. How can we communicate information reliably across a multi-hop network (as opposed to over just a single link or shared medium)?

A word on efficiency is in order. The techniques used to share the network and achieve reliability ultimately determine the *efficiency* of the communication network. In general, one can frame the efficiency question in several ways. One approach is to minimize the capital expenditure (hardware equipment, software, link costs) and operational expenses (people, rental costs) to build and run a network capable of meeting a set of requirements (such as number of connected devices, level of performance and reliability, etc.). Another approach is to maximize the bang for the buck for a given network by maximizing the amount of “useful work” that can be done over the network. One might measure the “useful work” by calculating the aggregate throughput (in “bits per second”, or at higher speeds, the more convenient "megabits per second") achieved by the different communications, the variation of that throughput among the set of nodes, and the average delay (often called the *latency*, measured usually in milliseconds) achieved by the data transfers. Largely speaking, we will be concerned with throughput and latency in this course, and not spend much time on the broader (but no less important) questions of cost.

Of late, another aspect of efficiency that has become important in many communication systems is *energy consumption*. This issue is important both in the context of massive systems such as large data centers and for mobile computing devices such as laptops and mobile phones. Improving the energy efficiency of these systems is an important problem.

■ 1.1.3 Scalability

In addition to reliability and efficient sharing, *scalability* (i.e., designing networks that scale to large sizes) is an important design consideration for communication networks. We will only touch on this issue, leaving most of it to later courses (6.033, 6.829).

■ 1.2 Outline and Plan

We have divided the course into four parts: the source, and the three important abstractions (signals, bits, and packets). For pedagogic reasons, we will study them in the order given below.

1. **The source.** Ultimately, all communication is about a source wishing to send some information in the form of messages to a receiver (or to multiple receivers). Hence, it makes sense to understand the mathematical basis for *information*, to understand how to *encode* the material to be sent, and for reasons of efficiency, to understand how best to *compress* our messages so that we can send as little data as possible but yet allow the receiver to decode our messages correctly. Chapters 2 and 3 describe the key ideas behind information, *entropy* (expectation of information), and *source coding*, which enables data compression. We will study Huffman codes and the Lempel-Ziv-Welch algorithm, two widely used methods.

2. **Bits.** The main issue we will deal with here is overcoming bit errors using error-correcting codes, specifically linear block codes and convolutional codes. These codes use interesting and somewhat sophisticated algorithms that cleverly apply redundancy to reduce or eliminate bit errors. We conclude this module with a discussion of the *capacity* of the binary symmetric channel, which is a useful and key abstraction for this part of the course.

3. **Signals.** The main issues we will deal with are how to modulate bits over signals and demodulate signals to recover bits, as well as understanding how distortions of signals by communication channels can be modeled using a *linear time-invariant* (LTI) abstraction. Topics include going between time-domain and frequency-domain representations of signals, the frequency content of signals, and the frequency response of channels and filters.

4. **Packets.** The main issues we will deal with are how to share a medium using a MAC protocol, routing in multi-hop networks, and reliable data transport protocols.