EECS563 Exam 3 Study Guide

Stephen Longofono

Fall 2017

The information herein was assembled from lecture notes and materials provided by Dr. James Sterbenz, along with book examples from Computer Networking: A Top-Down Approach by Kurose and Ross. This document is my interpretation of the materials presented in and outside the classroom for EECS 563 at the University of Kansas. Be sure to review the references for more authoritative information. [1][2][3][4]

1 "This will be on the exam..."

This section is comprised of things which were explicitly mentioned as exam fodder.

- Explain the difference between spectrum licensing and regulation.

  Licensed bands are sections of spectrum that require explicit permission to transmit on, usually in the form of an official license issued by a governing body like the FCC. Regulation of spectrum involves restrictions on how you can transmit, and generally apply to all bandwidths. For example, the 2.4 Ghz range is unrestricted and shared by many devices - it is unlicensed. However, it is still regulated. If you transmit with too much power, or too often, or in a way that jams other users, you will get into trouble.

- The last question on exam 2: "Sketch a block diagram for an IP router based on a fast packet switch architecture, labelling and explaining its blocks. You *must* show or use arrows to show the interconnection and relationship of the blocks. Describe the difference in functionality between a connection-oriented fast packet switch and a fast IP datagram router".

  See Figure 1. At the top, network signals including link costs are stored in a network topology database. These allow the router to run the link-state algorithm and fill in the forwarding table. Traffic signals are used by the routing processor to control its output queue scheduling and indicate to other routers that congestion is imminent. The routing processor maintains the longest-prefix forwarding table as a linear time means of directing incoming packets to appropriate ports on the switching fabric (in turn propagating them to the appropriate output queue for a given output interface). The forwarding table is used to program the switch fabric such that inputs are mapped to the correct outputs. At the inputs of the router, a link-layer de-encapsulation isolates the IP payload for use by the input processing unit. The input processing unit uses a shadow copy of the forwarding table to determine the appropriate input queue. At the outputs of the router, a link-layer encapsulation wraps the IP payload in a link layer frame with appropriate MAC addresses.

  The fast-packet IP switch has a number of differences from connection-oriented switches. Connection oriented switches go through a relatively longer connection setup, in which the route is determined and used to populate a Connection Identification (CID) table. The CID table is short and very easily processed, allowing neglibile latency for input processing. The fast packet IP switch introduced the longest-prefix matching and sophisticated switch fabric to keep up with connection oriented switches, without the overhead of setup and connection state.

- Why do we need both input and output queues? What problem does each solve?

  Input queues are necessary due to contention and HOL blocking. Contention is unavoidable; if two packets arrive which are destined for the same output interface, one of them has to go first and the other has to wait. The input queue allows the packet to wait. HOL blocking refers to a packet waiting due to contention blocking the packet that comes in behind it. If the packet behind it is destined for an idle output interface, the first packet at the "Head of the Line" is blocking the second. To get around this, the queue is monitored by some secheduling unit that can pull packets out of the queue when HOL blocking is detected.

  Output queues are used to react to network and link layer behaviors. If the switch fabric is much faster than the transmission speed, the output queue allows packets to buffer while transmission catches up. If traffic

signalling and in turn the output scheduler need to restrict a certain interface, the output queue allows packets to buffer.

- Describe channel partitioning and its purpose in networks

  Channel partitioning is one way to address medium access control (MAC). In wired or wireless networks, the channel used by the link layer to transmit data is limited if all parties use the same time domain, frequency, andor codes. If the channel can be partitioned such that each party has its own section of any of the above dimensions, then there are never collisions on that link. Channel partitioning is very effective if network load is well-defined: each party can be given only as many partitions as necessary, so everyone has the resources they need to transmit at the rate they wish. If the load is irregular, very high, or very low, this scheme is inefficient.

- What is the MAC scheme for antennae?

  Antennae can be directed, omnidirectional (monopole), or beamformed. If using an omnidirectional antenna, the signal is being broadcast in all directions, so some form of channel partitioning is required. For directed antennae (single or beamformed), the channel is partitioned by physical space such that parties will not interfere locally. There are still issues with hidden terminals to be addressed.

- Describe the difference between the 802.11 protocol and WiFi.

  802.11 is the protocol that defines how wireless communication is carried out in a wireless network. It specifies the initialization of connections, frame fields, means of handling channel contention, MAC, and general infrastructure. WiFi is a set of best practices and implementation details intended to help 802.11 devices interoperate. Much like the underwriter's laboratory specifies best practice implementations for inter-operability on the power grid, WiFi gives you reasonable assurance that your 802.11 device will be able to operate with any others without incident.

- Describe the improvement from a simple ALOHA MAC and one that uses time slots.

  ALOHA is a random access MAC, allowing any party to just begin talking in the channel whenever they have data to send. By Restricting this random talking to pre-determined time slots, collisions are either completely overlapping or not at all. That is, we have taken an infinite number of ways to interfere and reduced it to two case: complete interference or no interference. Statistically speaking, this has doubled the throughput.

- Describe the CSMACD MAC scheme and its extrema.

  Carrier-Sense Multiple Access MAC with Collision Detection depends on the ability to sense other parties trying to use the channel. When a party is ready to transmit, it listens for a set amount of time. If the line is in use, the party waits for a random exponential back-off period, and listens again. If the line is clear, the party begins to transmit, and regularly monitors for collision. If a collision is detected, the transmission is cancelled, and another random wait period is completed.

  In the worst case, another party could begin transmitting just before a packet arrives at its sensing unit; in this case, the collision would be detected after two propagation delays, one as the packet travels to the second party, and another as the collision propagates back to the sender. Thus as the propagation speed drops toward 0, the efficiency is 1, and as propagation speed grows large, efficiency goes to 0. From another point of view, as the transmission time for a maximum sized packet grows large, the efficiency approaches 1, since there is more time in which to detect a collision and stop transmission. The latter is not really useful, but it was in the slides so here it is.

- Describe the hidden terminal/node problem and how 802.11 solves it.

  The hidden node problem is related to how EM waves propagate in space - the signal power drops off as $\frac{1}{r^2}$. It is then possible for two transmitting parties to be far enough away that they cannot sense each other's transmissions, or for those parties to have an obstruction that hides their transmissions. If they are both trying to communicate with a third party somewhere between them (with a direct path to that third party), they could interfere with each other without being able to detect the interference.

  To address this problem, wireless networks use a form of Carrier Sense Multiple Access with Collision Avoidance (CSMACA). With EM waves, simultaneous transmitting and receiving is very difficult, so it is not feasible for a party to detect a collision as with CSMACD. Instead, a series of shorter messages is used to interact with other nodes. The sending party sends a short Request To Send (RTS) frame, and after a short time, the receiving party sends a Clear To Send (CTS) frame. Assuming these are omnidirectional antennae, any hidden terminals will also see the CTS frame, and if they are not waiting on it, they will wait until they hear an ACK indicating that the receiver is free again. More on this below.

- Describe the exposed node problem and one way to solve it.

  The exposed node problem happens when two receiving nodes are within range of one another. If a sender party is in range of a first party, but not a second party, it may cue the first party with a RTS, and the first party will respond with a CTS. Another sender in range of the second party, but not the first might send a RTS to the second party. The second party would not respond, since it is waiting to hear an ACK from the first party. In this way, two senders are interfering with each other indirectly; they are far enough apart to not physically interfere, but the exposed receivers are limiting communication.

  This problem can be solved by using directly antennae, such that any receiversaccess points are only broadcasting their CTS and ACK frames in a certain direction.

- This table

Table 1: MAC Characteristics

	Channel Partitioning	**Coordinated Access**	**Random Access**	**Spread Spectrum**
Types	TDMA, FDMA, WDMA, SDMA	token ring polling	ALOHA, slot CSMA/CD CA	CDMA, FH,DS
Complexity	low	high	low	high
Load Tolerance	high & deterministic	high	low	varies
Contention Yields	N/A	degraded performance	collisions	degraded performance
Resilience	poor	wired: OK wireless: poor	poor	resistant
Examples	802.11, 802.16	802.5	old ethernet 802.11	802.11

- Describe the MAC for 802.11.

  802.11 Uses three types of MAC: Spread spectrum limits the band in which all 802.11 devices operate, CSMACA avoids collision as a form of random access, and channel partitioning reduces interference via FDMA.

- Draw a diagram depicting two data frames being sent on a wireless channel using CSMACA.

  See Figure 2. When a party is ready to send, it listens to see if it can hear anyone using the channel. If it is clear, it waits for the Distributed Inter-Frame Spacing (DIFS) interval, and checks again. If the channel is still clear, it broadcasts a RTS frame. The receiver processes the RTS frame, and waits for the Short Inter-Frame Spacing (SIFS) interval. This shorter time is intended to make it harder to interrupt an exchange once it is underway. Later in the process, the sender waits the DIFS to allow others to speak if need be between data frames.

  After waiting the SIFS, the receiver broadcasts a CTS frame. This lets the sending party know it is OK to send data, and lets other would-be senders know that the channel is in use. Once the sender processes the CTS, it waits the SIFS interval (again, shorter to reduce likelihood of collision), and begins transmitting a longer data frame. Once the receiver has processed the data frame, it waits the SIFS again, and sends an acknowledgement frame.

  At this point, the sending party waits the DIFS interval, to allow other parties to interrupt if they need to send. If it is clear after the DIFS, the sender assumes it is OK and sends another frame. If another party were to send an RTS during the DIFS, the receiver would respond by a CTS, and the original sender would then need to start over after waiting.

  Note that at any given point, if a sender senses that the line is in use or sees a CTS it did not initiate, it uses the random exponential backoff intervals to wait just as with ethernet. Then, it waits the DIFS and tries again. This solves the same problem it did in ethernet: two senders collide, wait the same amount of time, and collide again indefinitely.

- Describe the problem of mobile nodes, and how mobile IP solves the problem.

  Mobile nodes are a problem for the Internet in general because every time a mobil endpoint moves to a different WLAN, the edge router for the new LAN would need to advertise the new IP location through it and the edge

router for the old LAN would need to cancel its advertisement for the IP address. This introduces large amounts of extra traffic to the network as millions of wireless devices move about.

Mobile IP works around the problem by establishing a home agent for mobile endpoints. The home agent is a relay for traffic to the mobile endpoint, and is responsible for maintaining a last known IP address to reach the mobile endpoint through.

When the mobile endpoint moves into a new network, it discovers a local entity (router) using ICMP. This local entity become the forein agent, and it calls home to the home agent to register the foreign agent. When some endpoint has IP datagrams for the mobile endpoint, it sends them to the home agent, which sends them to the mobile endpoint, "care of" the last known foreign agent. The mobile endpoint can reply directly to the sender, it does not use the home agent to relay its reply.

This creates cycling communications and has problems with latency in the handoffs as the mobile endpoint moves around. For the most part, the cellular network with data has replaced mobile IP, as the majority of truly mobile devices are smartphones. Also, for common mobile tasks like using your laptop to browse websites or email, a static IP is not neccesary - DHCP is sufficient.

- Describe the generations of wireless telephony (cell phones) and major characteristics.
  - 1G: The first generation was analog voice, built on top of the PSTN. Mobile switching centers were the gateway into the PSTN, and all signalling was analog.
  - 2G: Digital encoding of voice, adds hardware to support this. Digital voice means better quality when signal is strong, but drops off when signal is weak. The BSC is used to interact with towers, and passes traffic to a hierarchy of mobile switching centers which eventually reach the PSTN.
  - 2.5G: Adds in support for generalized radio packet service. This operated on the existing infrastructure for digital voice; since it was digital, there was no need to restrict network use to voice, but there were still limitations on rate because the network was designed to carry only voice.
  - 2.75G: Adds in EDGE, a precursor to data-optimized network. This improved upon existing infrastructure to better handle digital data. Technically meets the 3G standard depending on who you ask.
  - 3G: Adds in dedicated, data-optimized infrastructure to completely separate the paths of voice and data. The BSC is replace by a Radio Network Controller. The RNC sends voice traffic through a MSC hierarchy as before, and sends data traffic through a GRPS hierarchy that eventually attaches to the Internet. Eventually, Long Term Evolution further improves data rates, but does not quite meet the official standards of 4G.
  - 4G: The RNC is replaced by the eNodeB, because fuck you, that's why. Adds in an all-IP core network after traffic reaches the eNodeB. LTE-A achieves the data rates to qualitfy as 4G, and adds in more advanced partitioning and sharing of partitions at the radio leg of the network.
  - 5G: This future generation will need to support the Internet of Things, high data rates for both aggregate and individual use, direct P2P data, and low latency data for self-driving cars.

  See Table 2 for high-level overview.

- What is network jitter?

  Jitter refers to the variable delay associated with a network, in the context of multimedia packets arriving in a stream. Jitter varies with network load, which is why we need playback delays to buffer incoming data and absorb as much jitter as possible.

- Describe how a multimedia server manages traffic to its clients.

  A server needs to account for congestion and flow issues. For the former, the server throttles its output bandwidth to match the current throughput of the network. For the latter, the server, throttles its output bandwidth to match the rate at which the client can process incoming data (its input buffer).

- Describe the distinction between the service model of traffic management and the traffic management itself.

  The service model of traffic management is to facilitate the end-to-end transfer of layer 4 datagrams. The traffic management itself describe the actions taken at layer 3 and 4 to actually provide the service. The distinction is important because these two are at odds with each other; traffic management includes signalling over the network, which makes traffic worse.

Table 2: Cellular Generations

Generation	Digital/Analog	Service	Standard	Convergence?
1G	A	voice	AMPS & many more	not at all
2G	D	voice	GSM, CDMA	two major
2.5G	D	voice, piggybacked data	GSM, CDMA	two major
2.75G	D	voice, enhanced piggybacked data	GSM, CDMA	two major
3G	D	voice, dedicated data path	UTMS, CDMA2K	mostly UTMS toward LTE
3.9G	D	voice, dedicated data path, LTE rates, IP core	LTE	Mostly LTE
4G	D	Pure data, all-IP core, 4G speed	LTE-A	Mostly LTE-A

- What characteristics and parameters do we care about for traffic management?

  Traffic management systems are concerned with providing best-effort, probabilistic guarantees, or absolute guarantees for network throughput. That is, we try our best, we make it unlikely to see congestion, or we make damn sure that there will be no congestion. Together, they could be classified as QoS.

  The metrics/parameters we use to characterize traffic are:

  - Throughput - the rate we can push payloads through the network. We may subdivide this into peak rate, average rate, and 'burstiness'.
  - Delay - the end-to-end time required to deliver a payload
  - Jitter - the variability of our delay in the presence of traffic
  - Reliability - the amount of packets sent over the amount of packets delivered
  - Ordering - the order in which a sequence of packets arrives

- What is IntServ?

  Integrated Services (IntServ) is a signalling protocol for fine-grained traffic management on a per-flow basis. What is a flow? it is an ill-defined term. The RSVP protocol is defined to allow signalling and setup among routers to provide basic QoS. For well-defined expected loads, this allows resources to be reserved on routers in advance (dedicated queues) in an attempt to minimize the per-packet delay. It is not used in any backbone, and so it is not very useful.

- What is DiffServ?

  Differentiated Services (DiffServ) is a signalling protocol for coarse-grained traffic management on an aggregate basis. Aggregate flow through a network is managed statically by dedicating routes to certain sources and destinations. For example, the center of a major city with high traffic might be given 95% of available layer 3 resources, while the suburbs get only 5%. DiffServ makes use of the TOS field in IPV4, IPV6 has traffic class and flow fields built in.

- Compare and contrast QoS types

  The traditional Internet model is deliver most packets, most of the time, eventually. It makes no guarantees of delivery time, order, or integrity.

  Probabilistic QoS preemptively addresses traffic and uses signalling to reduce the likelihood of service interruption.

  Guaranteed QoS reserves resources in advance such that if a flow is active, it will always get the guaranteed throughput/delay/sequencing.

  Traditional Internet is purely best effort, and the majority of the modern Internet remains best-effort. Steps have been taken with router technology (fair queuing and weighted fair queuing) and signalling (ECN Explicit Congestion Notification) to increase overall performance, but there are ultimately no guarantees.

- Describe the mechanisms of ECN at layer 3 and 4

  For Explicit Congestion Notification, there are two sets of fields used. At layer 3, the TOS (Type of Service) field was hijacked to include two bits: 1 indicating that the given endpoint will support ECN, and another indicating that congestion was experienced along a route. If at any point along the route congestion is experienced, the CE bit is set high. When the reciever gets the message, it prepares a response with the layer 3 fields cleared.

  The response includes signals in the layer 4 control flags, indicating that the congestion window needs to be reduced. This is passed back to the sender, who them responds by adjusting the congestion window (recall that this self-throttles the output rate) and sends out a packet with the ECE (Explicit Congestion Notification Echo) control bit set, to let the receiver know it was changed.

- Why is traffic management hard?

  It is difficult to distinguish well-behaved traffic flows from misbehaving traffic flows. We can measure average performance over flows, but there are unpredictable periods of high traffic. Thus an attack and a period of unusually high traffic might be indistinguishable. The problem is made harder by the fact that ISPs are not consistent in their traffic engineering. IntServ is only effective if all ISPs use it. DiffServ cannot provide the fine-grained control to make QoS guarantees. IP in general is not aware of QoS. Finally, provisioning extra server resources is expensive, and generally ISPs will try to minimize their costs.

2 Major Topics

Physical Layer

- Types of line coding: binary vs. analog; amplitude, frequency, and phase

  Information can be encoded in a number of ways at layer 1, generally described by information theory.

  - Binary Codes - Binary codes rely on levels of analog voltage interpreted as digital information. They evolved to allow more information, to meet properties of physical media, and to allow clock synchronization. For early encodings, long intervals of 1 introduced DC bias that was problematic in hardware, and made synchronization difficult.
    - Unipolar has two levels: above and below a threshold. Thus each symbol has two possible values and two bits of information.
    - Bipolar has low, zero, and high. Thus each symbol has 3 possible values and three bits of information.
    - Differential encoding encodes a transition as a 0 and no transition as a 1.
    - Manchester encoding has two levels, and interprets a change from low to high as a 1 and a change from high to low as a 0. This solves the problems of synchronization since it must make some transition at each clock.
  - Analog Codes - Analog codes use frequency or amplitude modulation to encode digital information onto an analog signal. This allows for all of the above schemes for binary codes, along with new dimensions of information encoding.
    - Amplitude modulation multiplies a carrier signal by the bit value to be encoded. In the style of AM radio, many levels may be used to encode information.
    - Frequency modulation encodes information in the frequency of a constant-amplitude signal. Like FM radio, the gradations of available bandwidth encode the information.
    - Phase modulation encodes information by transmitting the same signal with a different phase offset. In combination with amplitude modulation, more sophisticated encodings can be used.
  - In general, some combination of the above are used, along with the physical properties of the media, to achieve the desired bandwidth of information.

- Types of physical media: wire, fiber, free space We transmit voltages through wire, optical through fiber, and EM waves in free space. We can further divide each category:

  - Un-shielded Twisted Pair - Copper wire pair. Simple, cheap, already installed in many homes. Reused PSTN lines saw 10 Mbps, CAT wire up to 100 Mbps. Examples: CAT5,6,7.
  - Shielded Twisted Pair - Same as above but with a conducting shield to protect from interference. Allows for much higher rates than un-shielded, but the added expense meant it was not widely adopted.

  - Coaxial cable - High-quality shielded cable used by cable companies. Great candidate for reuse of infrastructure, most homes are wired for this already.
  - Fiber - glass or plastic core with a protective sheath. Achieves much higher bandwidth due to speed of light. Problems with dispersion in the cheaper multimode version, so the single mode version is desirable. Expensive.
  - Free space - transmitting through the atmosphere with antennae. Tradeoff frequency for depth of information, ability to handle obstructions, atmospheric attenuation. Falls off as inverse of square, so limited distances in most cases. Includes microwave systems, cell towers, satellite, and radio.

- RF spectrum: ISM bands and licensing vs. regulation The spectrum we can use to communicate is heavily regulated and segmented. Certain bands are reserved for certain tasks - the entire spectrum is painstakingly divided. Many of these bands require exclusive licensing to use without penalty. All have regulations restricted what can be transmitted, at what power, and in what manner.

- Attenuation due to distance and frequency

  In all media, frequency and distance are limiting factors. For each type of media, this manifests itself in different ways. For wire, above a certain frequency, the signal will be attenuated. The signal is physically limited by unit impedance over long distances. For fiber, ignoring the mode, dispersion becomes problematic - the farther you go without a repeater, the further the wavefronts smear out. The index of refraction is a function of frequency, meaning that low frequency waves will travel farther and faster. Wireless communication attenuates very quickly because its power is equally distributed in space. It will fall off as the inverse squared of distance. For the same reasons as with optical media, dispersion is also problematic for higher frequencies. Wireless experiences atmospheric effects as well.

- Propagation: direct, reflection, diffraction, scattering, and multipath distortion

  Wireless media use is complicated by the absence of any shielding. A single, direct, point-to-point communication has to cope with several distinct kinds of interference. Reflection occurs when as sent signal reflects off an obstruction and eventually returns to the sender. Diffraction occurs when the wavelength is of similar length to an obstruction in the environment, and the wave bends around that obstruction. Scattering occurs when a signal hits a complicated surface and reflects in multiple and unpredictable ways. Multipath occurs when a signal reflects off multiple obstructions in a short period of time, which appears to the receiver as multiple time-shifted versions of the same signal. All of the above experience constructive and destructive interference, pushing attenuation to the inverse of the distance to the fourth power.

General Mobile Wireless Networks

- MAC, Wireless and Mobile Networking

  MAC for wireless and mobile networks is distinct from MAC addresses in wired systems. At the link layer, MAC was a protocol/algorithm for deciding who is who and when they can speak. In a wireless network, the problem of medium access is much more complicated for all the reasons in the physical layer stated above.

- Types of MAC

  Many of the MAC approaches for wired networks can be applied for wireless networks. The channel itself can be divided in time, frequency, space, or codespace as a channel partitioning scheme. The channel can be shared using a master/token indicator as with token rings or bus systems. The channel can be shared in a random way, with means of detecting interference and waiting for random amounts of time.

  - Channel partitioning: TDMA, FDMA, CDMA

    Channel partitioning can be done in any combination of time, frequency, space, and codespace. TDMA assigns time slots to certain parties, avoiding collisions by each only speaking on their turn. This is simple, but inefficient if there is asymmetric use or dead air. FDMA assigns frequency bands (within the larger spectrum bandwidth) to each party. These bands may be slightly overlapping, or orthogonal. CDMA assigns certain codes by which each communication is multiplied. By carefully choosing codes, multiple parties can speak simultaneously and the individual conversations can be discerned. Spatial partitioning is done with directed antennae, improving the attenuation at the same time by increased directivity. In general, channel partitioning is good for uniform loads and easy to implement, but struggles with variable load and is inefficient in low load.

- Coordinated MAC: token ring concepts

  Coordinated access is done via a master node cueing slave nodes to speak, or by a token which is passed from node to node granting access to the medium. This is the same as with token ring networks in the pre-ethernet days. In general, ring MAC is better at high or irregular load, but it is complicated to implement and easy to break if a wire gets cut.

- Random access: ALOHA, CSMA, CSMA/CD

  Random access means that each party speaks when it wishes. This can get messy if there is not some way to detect and recover from collisions. ALOHA was a simple random MAC with no rules. By combining this with time slots and forcing collisions to be all or nothing, the throughput was doubled. CSMA has each party listen for other speaking, and speaking if the channel is free. Various schemes were developed to address synchronization problems like exponential random backoff and random probabilistic transmit. CSMA/CD adds in the ability to detect if a collision is occurring and stop transmitting, improving the response to collisions. Random access is complicated and limits performance for high loads.

- Spread Spectrum

  Spread spectrum MAC - parties on the same channel use different codes so that they can talk at same time without interference. They may also do frequency hopping for security purposes and to further reduce the chance that they will interfere with each other.

  This is called DS CDMA for Direct Sequence Code Division Multiple Access. It entails multiplying all outbound data by a chipping code (which changes at a higher frequency relative to your bit rate), and sending that into the channel. Receivers multiply again by the code of the sender they want to listen to, and when they divide through by the length of the code, they recover the original message.

- Wireless network elements

  The elements of a wireless network are the base station, wireless nodes, and wireless links over which they communicate. For an 802.11 network, the base station is the AP (Access point), the nodes are end hosts with 802.11 capabilities, and the links are wireless. For a cell network, the base station is a BSC/RNC/eNodeB, the end hosts are wireless nodes, and the wireless links are direct and backhaul wireless connections.

- CSMA/CA and hidden nodes

  The hidden node problem can be solved by using CSMA/CA. There are two periods of time, DIFS and SIFS, which are used to coordinate access to the channel. When a sender is ready to talk, it listens to see if anyone else is using the channel. If not, it waits for the longer DIFS time, and listens again. If the channel is still free, then it sends a request to send frame (RTS) to the access point. The AP waits the shorter SIFS, then broadcasts a clear to send frame (CTS). All the senders that might use the AP will see the CTS, and if they did not request it, they will not talk until they hear an ACK. This is what prevents hidden nodes from interfering with each other. Once a node has the channel, it will wait for DIFS between each data frame to allow others to interrupt if they need to use the channel. In all cases, when the channel is sensed as busy, a random exponential backoff time is used, followed by a sense, DIFS, and another sense.

- 802.11 architecture and operation

  The general architecture of 802.11 networks is an AP which serves as the base station for any number of wireless node endpoints (phone, laptop, desktop, watch), which connects to a WLAN edge router (sometimes this is the access point), which is connected to the Internet via a wired LAN.

  There is also specification for Ad-Hoc mode, where the mobile endpoints communicate with each other and a router. In Ad-Hoc, the endpoints operate without access to the internet, so they have to implement DNS, MAC addresses, routing, and everything else that is done by layer 3 and 2 in an infrastructure network.

- 802.11 MAC control and frame flow (with IFS and RTS/CTS)

  802.11 Uses three types of MAC: Spread spectrum limits the band in which all 802.11 devices operate and uses CDMA to avoid interference from other bands, CSMACA avoids collision as a form of random access, and channel partitioning reduces interference via FDMA. Frame flow is as described above for CSMA/CA

- Bluetooth/802.15 and 802.16 high-level concepts 802.16 Defines WMAN Wireless Metropolitan Area Networks and WiMax. The idea is to have longer range BSS areas in highly populated regions with sufficient bandwidth for all users. The architecture is similar, with a hierarchy of access points to a high-bandwidth network.

Not widely used and not likely to be in the future. WiMAX was a set of specifications for the architecture, component roles, and service standards for WMAN.

Bluetooth was originally a proprietary standard that has since been adopted for WPAN (wireless personal area networks). The original standard has tiny address space and issues with throughput since all communication was relayed through a master node. Once it came into the public domain as 802.15, it was further developed to allow p2p communication, have a larger address space, and can operate in very low power (BLE). 802.15 uses frequency hopping and TDMA to partition the channel.

802.15 also includes zigbee and sensor networks. Zigbee is for low throughput, small area networks that operate on very low power. Sensors networks have similar qualities, but are often designed to relay information and be robust to failure.

- Mobility and MobileIP

  Mobility in wireless networks refers to how a node can move among networks, within the same network, or outside of local networks and still maintain connectivity. A fixed wireless node might be your desktop computer, a mobile node might be your smart watch. A nomadic node might be your smartphone.

  Mobile IP is required when you need have a fixed IP address where others can always reach you, but you are moving among different networks. Without mobile IP, you would need to advertise your IP at every new network you join, and let every other router on the Internet know where you are now every time you move (not feasible). Instead, mobile IP lets you register the nearest router's IP as your "foreign agent". You have a home network that takes all traffic for your fixed IP, and forwards it to the foreign agent for you. This lets you move around to different networks, and as long as you register foreign agents to your home network, there is a path to your end host using the fixed IP address.

  The cell phone network has mostly replaced the need for this. In most cases, DHCP is sufficient from Internet use in remote networks; you don't need a fixed IP to browse websites, use email, etc.

- Mobile telephone generation major characteristics

  - 1G: The first generation was analog voice, built on top of the PSTN. Mobile switching centers were the gateway into the PSTN, and all signalling was analog.
  - 2G: Digital encoding of voice, adds hardware to support this. Digital voice means better quality when signal is strong, but drops off when signal is weak. The BSC is used to interact with towers, and passes traffic to a hierarchy of mobile switching centers which eventually reach the PSTN.
  - 2.5G: Adds in support for generalized radio packet service. This operated on the existing infrastructure for digital voice; since it was digital, there was no need to restrict network use to voice, but there were still limitations on rate because the network was designed to carry only voice.
  - 2.75G: Adds in EDGE, a precursor to data-optimized network. This improved upon existing infrastructure to better handle digital data. Technically meets the 3G standard depending on who you ask.
  - 3G: Adds in dedicated, data-optimized infrastructure to completely separate the paths of voice and data. The BSC is replace by a Radio Network Controller. The RNC sends voice traffic through a MSC hierarchy as before, and sends data traffic through a GRPS hierarchy that eventually attaches to the Internet. Eventually, Long Term Evolution further improves data rates, but does not quite meet the official standards of 4G.
  - 4G: The RNC is replaced by the eNodeB, because fuck you, that's why. Adds in an all-IP core network after traffic reaches the eNodeB. LTE-A achieves the data rates to qualify as 4G, and adds in more advanced partitioning and sharing of partitions at the radio leg of the network.
  - 5G: This future generation will need to support the Internet of Things, high data rates for both aggregate and individual use, direct P2P data, and low latency data for self-driving cars.

  See Table 2 for high-level overview.

Security and Resilience

- Communication and threat model

  For communications, there are a number of functions we might want for security reasons:

- Confidentiality - we may only want the message contents visible to sender & the intended receiver
- Authentication - we may want to confirm identity of someone we are talking to
- Message integrity – we may want to guarantee that message is not altered without detection
- Nonrepudiation – We may want to make sure the sender of a message can't deny sending it
- Access control and availability of resources - We may want to restrict access to resources and services to legitimate and authorized users

We organize threats in terms of the above; this helps us address attacks specifically base on what we care about. The slides say this is important because we always need to justify why we are taking the security steps. Some common attacks include:

- Eavesdropping - This is a threat to confidentiality. Attackers may try to listen to authentication messages or extract data from our messages.
- Authenticity - These threats try to compromise authentication by spoofing, impersonation, forgery, and identity theft to access secure data, resources, money, etc.
- Integrity - These threats alter or remove messages to interrupt or remove communication.
- Authentication & resource availability - These threats attempt to disrupt authentication or normal service by DDOS attacks and the like. Attacks may also repeat messages that have been sent in the past (replay attack).

- Security functions and services: confidentiality, integrity, digital signatures, authentication
  - Confidentiality - We address threats to confidentiality by using cryptography. In order to be properly and completely implemented, it must occur end-to-end, and for multiuser systems, application to application.
  - Integrity & Digital Signatures - Integrity attacks try to alter message contents for various reasons. To handle this, networks use digital signatures and authentication hashes. An example of authentication hashes is having all senders and receivers use a cryptographic hash function and a secret key. All messages are sent in plaintext along with a hash of the same message plus a secret key (say the secret key plus the timestamp). At the receiver end, the plaintext message and secret key are fed into the hash, and if it doesn't match the hash sent with the message, it has been tampered with and it gets dropped. Digital signatures achieve the same end by using public key cryptography. A hash of the message by itself is encrypted with the public key of the receiver, and appended to the plaintext message. At the receiver end, the hash of the message is decrypted using their private key, then they compare it to the plaintext message run through the hash. If they do not match, then the message has been tampered with. Note that cryptographic hashes are generally not easily invertible.
  - Authentication - Authentication is tied to the above integrity functions through the digital signature. For session identification and to stop replay attacks, a timestamp and a unique, one time use value called a nonce will be put through the hash along with message. This prevents attackers from replaying messages exchanged during an authentication handshake, as the nonce value will no longer be valid. In general, the nonce value is tied to the authentication key for that session.

    This does not stop an attacker from impersonation, as the public key and hash are available. For this reason, additional passwords, a certificate authority, and 2-factor authentication are needed to improve security.

- Public key versus symmetric key cryptography

  Public key cryptography has pairs of public and private keys assigned to each party. The public key is shared and available to everyone, and private keys are kept secret. When you want to send to message to someone, you use their public key with the RSA algorithm to encrypt the message, and then it can only be decrypted with the matching private key. This is more secure than other cryptography but it is much more expensive. Typically, you use public key cryptography to sign messages and exchange symmetric keys, and then use symmetric keys to encrypt everything else.

  Symmetric keys like AES, DES, 3DES, are complicated block-cipher algorithms for encryption. So long as both parties know the keys, the data can be quickly encrypted and decrypted, and it is difficult for attackers to crack them. Public key is still necessary since you still need to get the symmetric key to the reciever securely in order to use it.

- cryptography concepts; DES and AES high-level features

  DES (Data Encryption standard) was one of the first symmetric key schemes, using 56 bit key length in a block-cipher. Data segements of 64 bits are run through table mappings (one for each of the 56 bits of the key), and this is repeated 7 times to use all 56 key tables. This can be broken by modern computers in minutes. As a stopgap solution, 3DES was used, running everything through DES three times to make it harder to break.

  AES (advanced encryption standard) improved upon DES by expanding to 128-1920 bit keys, with more sophisticated swapping of intermediate results and a highly parallelizable algorithm. This is the modern standard, and would take billions of years to crack with today's computers.

- certificates and revocation Certificate Authorities are verification agencies. They are necessary because we need a way to associate people with sets of public encryption keys. When a person provides adequate identification, their key pair is associated with them. To verify that someone is who they say they are, the certificate authority uses its private key to digitally sign the public key of the person they are certifying (along with other information about when it was issued, when it expires, etc.). The person sends their certificate as a part of authentication, and the certificate authority will issue a digital signature of that certificate for comparison.

  Certificates are often ignored, or expired, so they are not as effective as they should be. They may be revoked if they expire, or if the authority has reason to believe the certificate has been compromised.

- end-system protection: virus scanners, firewalls, and IDSs Virus scanners periodically scan the hard disk and the registry for known viruses and programs that have a similar structure to known viruses. They can be effective against known attack vectors, but new worms and viruses will likely pass undetected.

  Firewalls allow a network to observe and filter packets moving through the network. Filtering firewalls simply matches a threat signature and drops any packets that match. Firewalls are employed in enterprise and personal settings, since an enterprise firewall cannot prevent intrusions within the network.

  Stateful firewalls maintain records of open connections, to allow packets that would otherwise be rejected to pass through if they were part of a known connection. This allows for more sophisticated filtering that is generally less disruptive to users.

  Intrusion detection systems are placed in various places throughout a network to do "deep packet inspection". This entails correlating packets in time, looking for known attack signatures at the individual and aggregate level. Often multiple layers are used to allow this to happen at speed without interrupting flow.

- application security motivation; HTTPS, SSH, and secure email concepts Application security is necessary in addition to all other security because we have no guarantee that the entire path is secure (security can only be correctly implemented end-to-end).

  Internet browsing uses HTTPS, which is essentially a means of conducting HTTP over SSL/TLS. The entire HTTP header and payload are encrypted, preventing access to everything above layer 4.

  Email encryption is implemented using PGP, which is loosely based on digital signatures and public/symmetric keys. Email security depends on the mode of access - HTTPS is used for html-based email clients, SSL/TLS is used for POP3 and IMAP access.

  Other applications like SSH allow terminal access to remote machines via a specialty security protocol.

References

[1] J. Kurose and K. Ross. *Computer Networking: A Top-Down Approach, 7th Ed.* Pearson, London, 2017.

[2] J. Sterbenz. *Introduction to Communication Networks: Mobile and Wireless Networks.* URL: `https://www.ittc.ku.edu/~jpgs/courses/intronets/lecture-mobilewireless-intronets-display.pdf`.

[3] J. Sterbenz. *Introduction to Communication Networks: Multimedia, Traffic, and Session Control.* URL: `https://www.ittc.ku.edu/~jpgs/courses/intronets/lecture-mm-tm-sess-intronets-display.pdf`.

[4] J. Sterbenz. *Introduction to Communication Networks: Security, Resilience, and Reliability.* URL: `https://www.ittc.ku.edu/~jpgs/courses/intronets/lecture-security-resilience-intronets-display.pdf`.

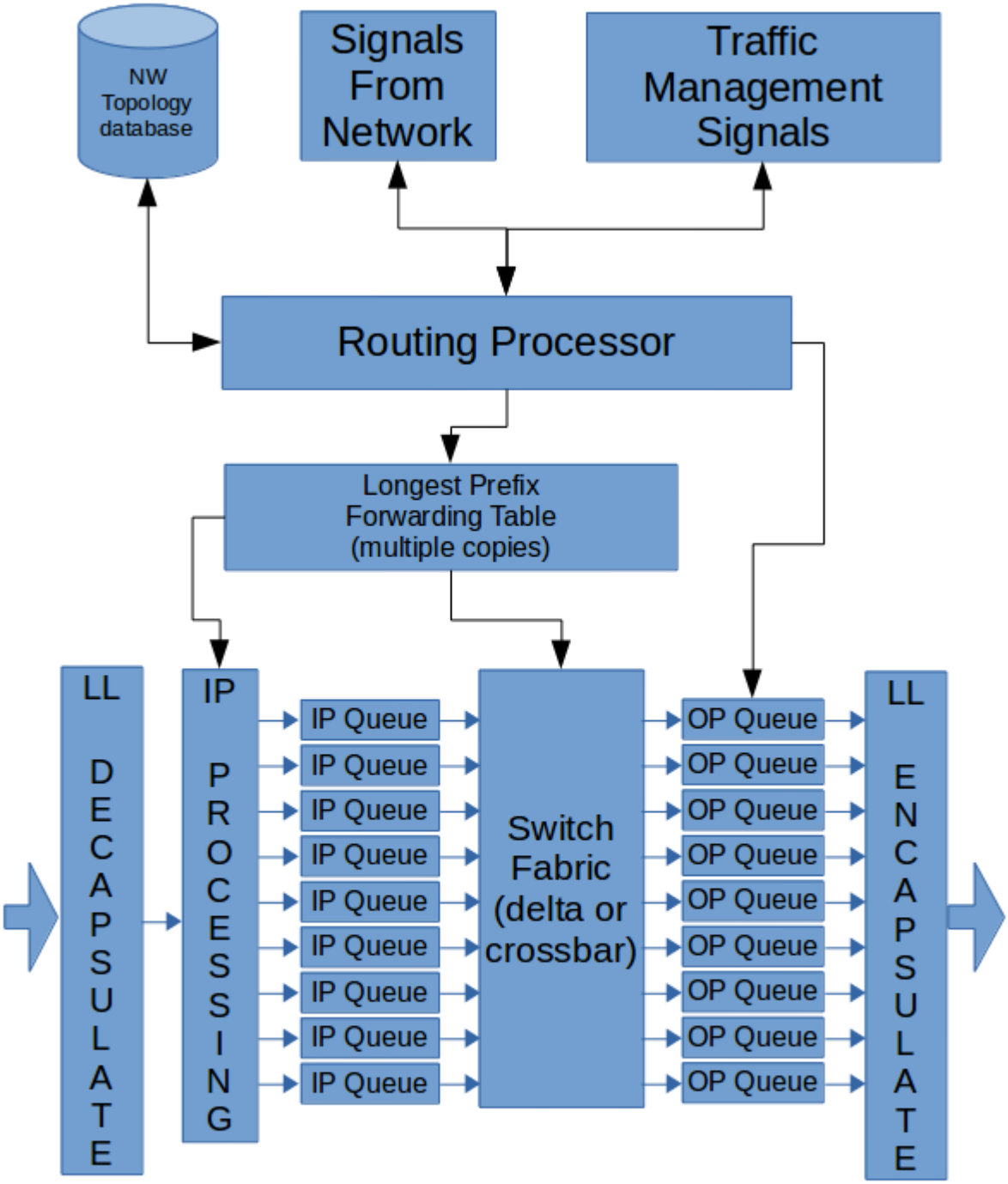

Figure 1: IP Fast Packet Switch Architecture

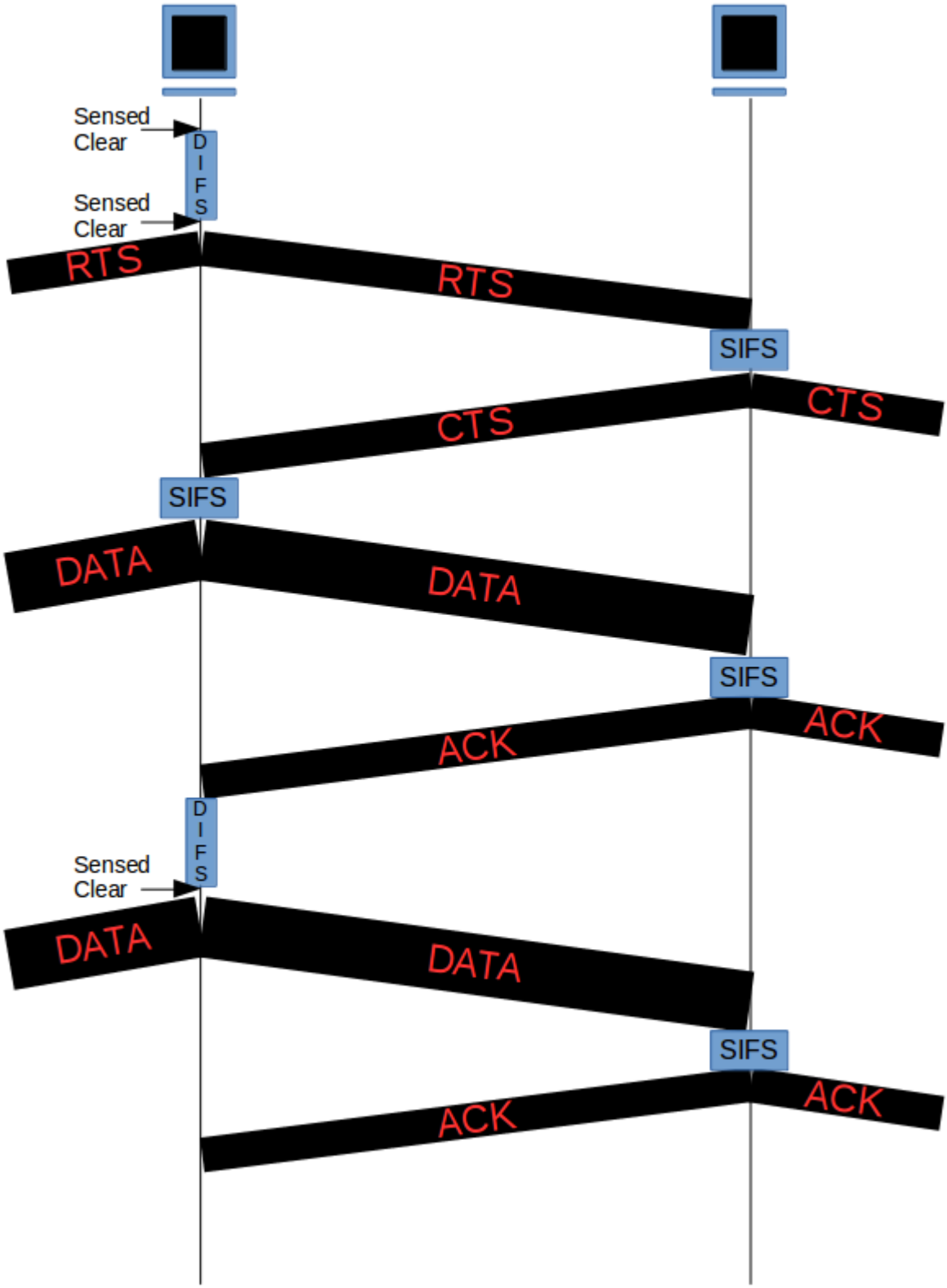

Figure 2: Diagram of CSMA/CA Data Exchange