- Chapter 3 Link Layer
- 3.1 Introduction
  - Purpose of the link layer in the TCP/IP protocol suite is to send and receive IP datagrams for the IP module
    - it is used to carry other protocols that help support IP
  - TCP/IP supports many different link layers, depending on the type of networking hardware being used: wired LANs such as Ethernet, metropolitan area networks (MANs)
  - Tunneling - link layer protocols can be carried inside other protocols
  - Frame - term used for referring to a link layer PDU
  - frame formats usually support a variable-length frame size ranging from a few bytes to a few kilobytes
    - maximum transmission unit (MTU) - the upper bound of the range
- 3.2 Ethernet and the IEEE 802 LAN/MAN Standards
  - Example
    - a basic shared ethernet network consists of one or more stations attached to a shared cable segment. Link-layer PDUs (frames) can sent from one station to one or more others when the medium is determined to be free. If multiple stations send at the same time, possibly because of signal propagation delays a collision occurs. Collisions can be detected, and they cause sending stations to wait a random amount of time before retrying. This common scheme is called carrier sense, multiple access with collision detection.
  - because multiple stations share the same network, this standard includes a distributed algorithm implemented in each Ethernet network interface that controls when a station gets to send data it has.
  - carrier sense, multiple access with collision detection (CSMA/CD), mediates which computers can access the shared medium (cable) without any other special agreement or synchronization
  - CSMA/CD a station first looks for a single currently being sent on the network and sends its own frame when network is free
    - this is carrier sense portion of the protocol
    - if some other station happens to send at the same time, the resulting overlapping electrical signal is detected as a collision
    - in this case each station waits a random amount of time before trying again
      - the amount of time is selected by drawing from a uniform probability distribution that doubles in length each time a subsequent collision is detected
    - eventually, each station gets its chance to send or times out trying after some number of attempts
  - CSMA/CD only one frame is traveling on the network at a given time
  - Methods such as CSMA/CD are more formally called Media Access Control (MAC) protocols
    - many types of MAC protocols
      - some are based on having each station try to use network independently (contention based protocols likes CSMA/CD)
      - others are based on prearranged coordination (e.g. by allocating time slots for each station to sent)
  - A switched Ethernet network consists of one or more stations, each of which is attached to a switch port using a dedicated wiring path. In most cases where switched Ethernet is used, the network operates in a fill-duplex fashion and the CSMA/CD algorithm is not

required. Switches may be cascaded to form larger Ethernet LANs by interconnecting switch ports, sometimes called “uplink” ports.
  - full-duplex Ethernet - the ability for a station to send and receive data at the simultaneously
  - half-duplex - one direction at a time
  - most of the capabilities for TCP/IP for Ethernet networks are also used for wi-fi networks
- 3.2.1 The IEEE 802 LAN/MAN Standards
  - IEEE standards with the prefix 802 define the operations of LANs and MANs
  - Other than the specific types of LAN networks defined by the 802.2, 802.11, and 802.16 standards, there are some related standards that apply across all of the IEEE standard LAN technologies
    - common to all three of these is the 802.2 standard
      - Logical Link Control (LLC) frame header common among many of the 802 networks’ frame formats. In IEEE terminology, LLC and MAC are “sublayers” of the link layer, where the LLC (mostly frame format) is generally common to each type of network and the MAC layer may be somewhat different
      - Original Ethernet made use of CSMA/CD, WLANs often make use of CSMA/CA (CA is “collision avoidance”)
- 3.2.2 The Ethernet Frame Format
  - All Ethernet (802.3) frames are based on a common format
  - Ethernet frame
  - Preamble area - used by the receiving interface’s circuitry to determine when a frame is arriving and to determine the amount of time between encoded bits (called clock recovery)
  - Ethernet is an asynchronous LAN (i.e. precisely synchronized clocks are not maintained in each Ethernet interface card) the space between encoded bits may differ somewhat from one interface card to the next
  - The preamble is a recognizable pattern, which the receiver can use to “recover the clock” by the time the start frame delimiter (SFD) is found. The SFD has the fixed value 0xAB
  - Example
    - The Ethernet frame format contains source and destination addresses, an overloaded Length/Type field, a field for data, and a frame check sequence (a CRC32). Additions to the basic frame format provide for a tag containing a VLAN ID and priority information and more recently for an extensible number of tags. The preamble and SFD are used for synchronizing receivers. When half-duplex operation is used with Ethernet running at 100Mb/s or more, additional bits may be appended to short frames as carrier extension to ensure that the collision detection circuitry operates properly
  - Basic format
    - 48-bit Destination (DST) and Source (SRC) Address fields
      - sometimes known by other names such as “MAC address”, “link-layer address”, “802 address”, “hardware address,” or “physical address.”
      - destination address in an Ethernet frame is also allowed to address more than one station (called “broadcast “ or “multicast”)
      - The broadcast capability is used by the ARP protocol and multicast capability is used by the ICMPv6 protocol to convert between network-layer and link-layer addresses
    - Type field that doubles as a length field
      - identifies the type of data that follows the header
      - popular values with TCP/IP networks include IPv4(0x0800),IPv6(0x86DD), and ARP(0x0806)
    - payload portion of the frame or data area

- area where higher-layer PDUs such as IP datagrams are placed Traditionally, the payload area for Ethernet has always been 1500 bytes, representing the MTU for Ethernet
- the payload is sometimes padded with 0 bytes to ensure that the overall frame meets the minimum length requirements

3.2.2.1

- Cyclic Redundancy Check (CRC)
  - provides an integrity check on the frame
  - includes 32 bits and is sometimes known as the IEEE/ANSI
  - to use an n-bit CRC for detection of data transmission in error, the message to be checked is first appended with n 0 bits, forming the augmented message. Then, the augmented message is divided (using modulo-2 division) by an (n + 1) - bit value called the generator polynomial, which acts as the divisor
  - the value that is placed in the CRC field of the message is the one's complement of the remainder of this division. Generator polynomials are standardized for a number of different values of n.
  - Frame Check Sequence (FCS) or CRC contains the result of long modulo-2 division
  - Upon receipt, the receiver performs the same division and checks whether the value in the FCS field matches the computed remainder
    - If the two do not match the frame was likely damaged in transit and is discarded. The CRC family of functions can be used to provide a strong indicator of corrupted messages because any change in the bit pattern is highly likely to cause a change in the remainder term.

- 3.2.2.2 Frame sizes
  - minimum size of Ethernet frames is 64 bytes, requiring a minimum data area (payload) length of 48 bytes (no tags)
    - In cases where the payload is smaller, pad bytes (value 0) are appended to the end of the payload portion to ensure that the minimum length is enforced
  - maximum frame size of conventional Ethernet is 1518 bytes (including the 4-byte CRC and 14-byte header).
    - if the frame contains an error (detected on receipt by an incorrect CRC) only 1.5KB need to be retransmitted to repair the problem
    - the size limit MTU to not more than 1500 bytes
    - in order to send a larger message multiple frames are required (64KB a common larger size used with TCP/IP networks, would require at least 44 frames)
    - each frame contributes a fixed overhead
    - inter-packet-gap (IPG) - 12 byte times
- 3.2.3 802.1p/q: Virtual LANs and QoS Tagging
  - virtual LANs - Compliant Ethernet switches isolate traffic among hosts to common VLANs.
    - Note that because of this isolation, two hosts attached to the same switch but operating on different VLAN's require a router between them for traffic to flow.
  - station-to-VLAN mapping
    - assigning VLANs by port is a simple and common method, whereby the switch port to which the station is attached is assigned a particular VLAN, so any station so attached becomes a member of the associated VLAN.
    - MAC-address-based VLANs that use tables within Ethernet switches to map a stations MAC address to corresponding VLAN.
  - trunking - when multiple v lans must span multiple networks
  - VLAN tag - holds 12 bits and the VLAN identifier used when van spans multiple switches

- contains 3 bits for QOS support
  - 802.1p standard
  - administrator must configure the ports
- some switches support
  - default value native VLAN values
- 802.1p
  - contains a mechanism to express a QoS identifier on each frame
  - includes a 3 bit priority field
    - eight classes of service are defined
      - class 0 the lowest priority is for conventional bess traffic
      - class 7 is the highest priority and might be used for critical routing or network management functions
    - standards specify how priorities are encoded in packets but leave the policy that governs which packets should receive which class and the underlying mechanisms implementing prioritized services, to be defined by the implementer
- vcon-fig - linux command for manipulating 802.1p priorities
  - uses
    - add or remove virtual interfaces associating VLAN IDs to physical interfaces
    - set 802.1p priorities
    - change the way virtual interfaces are identified
    - influence the mapping between packets tagged with certain VLAN IDs and how they are prioritized during protocol processing in the operating system
- linux virtual interface commands
  - default method of naming virtual interfaces in Linux is based on concatenating the associated physical interface with the VLAN ID
    - VLAN ID 2 associated with the interface ethic is called eth1.2

- 3.2.4 802.1AX: Link Aggregation (Formerly 802.3ad)
  - Link aggregation - two or more interfaces are treated as one in order to achieve greater reliability through redundancy or greater performance by splitting (striping) data across multiple interfaces.
  - Link Aggregation Control Protocol (LACP) to manage such links
  - Link aggregation on Ethernet switches that support it can be a cost effective alternative to investing in switches with high-speed network ports.
  - Link aggregation may be supported not only on network switches but across multiple network interface cards (NICs) on a host computer
  - often aggregated ports must be of the same type
  - Linux link aggregation
    - bonding command - device driver supporting link aggregation
    - ifenslave command -
      - master and slave labeling
  - Example
    - bond0 is assigned the IPv4 address information we would typically assign to either of the individual interfaces
      - receives the first slave mac address by default
    - When IPv4 traffic is sent out of the bond0 virtual interface options in which slave will carry the information. The options are selected using arguments provided when the bonding driver is loaded. Mode option determines the method.
      - round-robin deliver is used between the interfaces
      - one interface acts as a backup to the other

        - used for high-availability systems that can fail over to a redundant network infrastructure if one link has ceased functioning
      - the interface is selected based on performing XOR of the MAC source destination addresses
        - is intended to choose the slave interface based on the traffic flow
        - with enough different destinations, traffic between the two stations is pinned to one interface
      - frames are copied to all interfaces
        - for fault tolerance
      - 802.3ad standard link aggregation is performed
        - used with 802.3ad capable switches, to enable dynamic aggregation over homogeneous links
      - more advance load-balancing options are used
  - LACP
    - designed to make the job of setting up link aggregation simpler by avoiding manual configuration
    - Example
      - LACP “actor” (client) and “partner” (server) send LACPDUs every second once enabled
      - LACP automatically determines which member links can be aggregated into a link aggregation group (LAG) and aggregates them
        - accomplished by sending a collection of information across the link
      - a receiving station can compare the values it sees from other ports and perform the aggregation if they match
- 3.3 Full Duplex, Power Save, Autonegotiation, and 802.1X Flow Control
  - Ethernet
    - operating using a shared cable
    - data could be sent only one way at a time (half duplex)
  - Switched Ethernet
    - sets of many links
    - multiple pairs of stations could exchange data simultaneously
    - modified to operate in full duplex
      - effectively disabling the collision detection circuitry
      - allowed the physical length of the Ethernet to be extended, because the timing constraints associated with half-duplex operation and collision detection were removed
  - Linux commands
    - ethtool
      - program can be used to query whether full duplex is supported and whether it is being used
      - can also display and set many other interesting properties of en Ethernet interface
  - autonegotiation - mechanism originating with 802.3u to enable interfaces to exchange information as such as speed and capabilities such as half- or full-duplex operation.
    - autonegotiation information is exchanged at the physical layer using signals sent when data is not being transmitted or received
  - Port, PHYAD, Transceiver -
    - identify the physical port type,
    - its address
    - whether the physical-layer circuitry is internal or external to the NIC

- current message-level - value is used to configure log messages associated with operating modes of the interface
  - its behavior is specific to the driver being used

- 3.3.1 Duplex Mismatch
  - duplex mismatch - when a computer and its associated switch port are configured using different duplex configurations or when autonegotiation is disabled at one end of the link but not the other
    - connection does not completely fail but instead may suffer significant performance degradation
- 3.3.2 Wake-on LAN (WoL), Power Saving, and Magic Packets
  - Wake-Up Capabilities (Windows) and Wake-On (Linux) - options are used to bring the network interface and/or host computer out of a lower-power (sleep) state based on the arrival of certain kinds of packets
  - Wake-On (Linux) - values are zero or more bits indicating whether receiving the following types of frames trigger a wake-up from a low-power state
    - physical-layer (PHY) activity (p)
    - unicast frames destined for the station (u)
    - multicast frames (m)
    - broadcast frames (b)
    - ARP frames (a)
    - magic packet frames (g)
      - and magic packet frames including a password
    - example
      - ethtool -s eth0 wol umgb
  - Magic packet frames
    - contain a special repeated pattern of the byte value 0xFF
    - often such frames are sent as a form of UDP packet encapsulated in a broadcast Ethernet frame
    - Example
      - wol 00:08:74:93:C8:3C
      - command is to construct a magic packet
  - Wireshark
    - magic packet frame in Wireshark begins with 6 0xFF and then repeats the MAC address 16 times
- 3.3.3 Link-Layer Flow Control
  - flow control (slow senders down)
    - implemented by sending special signal frames between switches and NICs
    - signals the sender that it must slow down its transmission rate, although the specification leaves the details of this to the implementation
    - Ethernet - uses an implementation of flow control called PAUSE messages also called PAUSE frames
      - PAUSE frames are always sent to the MAC address 01:80:C2:00:00:01 and are used only on full-duplex links
      - they include a hold-off time value indicating how long the sender should pause before continuing to transmit
      - use a MAC control frame - is a frame format using the regular encapsulation, but with 2 byte opcode immediately following the Length/Type field
        - pause frames use a MAC control frame but with a 2-byte quantity encoding the hold-off time

- Bridges and Switches
  - switches - high performance bridges
  - a bridge or switch is used to join multiple physical link-layer networks or groups of stations
  - Example
    - a simple extended Ethernet LAN with two switches. Each switch port has a number for reference, and each station (including each switch) has its own MAC address
    - ports are numbered for reference
    - every network element, including each switch, has its own MAC address
  - Nonlocal MAC addresses are “learned” by each bridge over time so that eventually every switch knows the port upon which every station can be reached
    - these lists are stored in tables (called filtering databases) within each switch on a per-port (and possibly per-VLAN) basis
  - a standard computer with multiple interfaces can be used as a bridge
- bridge control linux
  - brct1 addbr br0
    - creates the bridge
  - brct1 addif br0 eth0
    - adds the interfaces to the bridge
    - delif removes interfaces
  - showmacs
    - can be used to inspect the filter databases (called forwarding databases or fobs in Linus terminology)
- Each time an address is learned, a timer is started
  - in Linux, a fixed amount of time associated with the bridge is applied to each learned entry the entry is removed as indicated here
  - When an entry is removed because of aging, subsequent frames for the removed destination are once again sent out of every port except the receiving one (called flooding), and the entry is placed anew into the filtering database
- If the tables are empty the network experiences more overhead but still functions
- 3.4.1 Spanning Tree Protocol (STP)
  - Bridges may operate in isolation, or in combination with other bridges
  - Two or more bridges are in use the possibility exists for a cascading, looping set of frames to be formed
  - works by disabling certain ports at each bridge so that topological loops are avoided, yet the topology is not partitioned—all stations can be reached
  - mathematically, a spanning tree is a collection of all of the nodes and some of the edges of a graph such that there is a path or route from any node to any other node (spanning the graph), but there are no loops (the edge set forms a tree).
  - There can be many spanning trees on a graph
    - STP finds one of them for the graph formed by bridges as nodes and links as edges
- STP Example
  - Using STP, the B-A, A-C, and C-D links have become active on the spanning tree. Ports …. are in the forwarding state; all other ports are blocked. This keeps frames from looping and avoids broadcast storms. If a configuration change occurs or a switch fails, the blocked ports are changed to the forwarding state and the bridges compute a new spanning tree.
  - frames are created only as a result of another frame arriving. There is no amplification
- Bridge Protocol Data Units
- 3.4.1.1 Port States and Roles
- Each port in each bridge may be in one of five states

  - blocking, listening, learning, forwarding, and disabled
- Blocking
  - after initialization a port enters the blocking state
  - does not learn addresses, forward frames, or transmit BPDUs
  - it does monitor received BPDUs in case it needs to be included in the future on a path to the root bridge, in which case it needs to be included in the future on a path to the root bridge, in which case the port transitions to the listening state
- Listening state
  - the port is now permitted to send as well as receive BPDUs but not learn addresses or forward data
- Learning state
  - a port enters the leaning state after a typical forwarding delay timeout of 15s
  - permitted to do all procedures except forward data
  - it waits another forwarding delay before entering the forwarding state and commencing to forward frames
- Port State Machine (PSM)
  - each port is said to have a role
  - root port, designated port, alternative port, or backup port
  - root port - ports at the end of an edge on the spanning tree headed toward the root
  - designated ports - are ports in the forwarding state acting as the port on the least-cost path to the root from the attached segment
  - alternative port - are other ports on an attached segment that could also reach the root but at higher cost
    - not in the forwarding state
  - backup port - is a port connected to the same segment as a designated port on the same bridge
    - could easily take over for a failing designated port without disrupting any of the rest of the spanning tree topology but do not offer an alternate path to the root should the entire bridge fail.
- 3.4.1.2 BPDU Structure
- BPDUs are always sent to the group address 01:80:C2:00:00:00 and are not forwarded through a bridge without modification
- Format figure 3-15
  - BPDUs are carried in the payload area of 802 frames and exchanged between bridges to establish the spanning tree. Important fields include the source, root node, cost to root, and topology change indications
- The Protocol (Prot) field gives the protocol ID number, set to 0
- The Version (Vers) field is set to 0 or 2, depending on whether STP or RSTP is in use
- Type field assigned similarly
- Flags field contains Topology Change (TC) and Topology Change Acknowledgment (TCA) bits
- Root ID field - gives the identifier of the root bridge in the eyes of the sender of the frame, whose MAC address is given in the Bridge ID field
- root path cost - is the computed cost to reach the bridge specified in the Root ID field
- PID field is the port identifier and gives the number of the port from which the frame was sent appended to a 1 byte configurable Priority field
- Message A field gives the message age
  - not a fixed value like other time-related fields
  - when the root bridge sends a BPDU it sets this field to 0

  - any bridge receiving the frame emits frames on its non-root ports with the Message Age field incremented by 1
  - the field acts as a hop count, giving the number of bridges by which the BPDU has been processed before being received
  - when a BPDU is received on a port, the information it contains is kept in memory and participates in the STP algorithm until it is timed out, which happens at time (MaxA - MsgA)
    - should time pass on a root port without receipt of another BPDU, the root bridge is declared “dead” and the bridge starts the root bridge election process over again
- Maximum Age field gives the maximum age before timeout
- Hello Time field gives the time between periodic transmissions of configuration frames
- Forward Delay field gives the time spent in the learning and listening states
- 3.4.1.3 Building the Spanning Tree
- First job of STP is to elect the root bridge
  - the bridge in the network (or VLAN) with the smallest identifier
- when a bridge initializes, it assumes itself to be the root bridge and sends configuration BPDUs with the Root ID field matching its own bridge ID
  - if it detects a bridge with a smaller ID, it ceases sending its own frames and instead adopts the frame it received containing the smaller ID to be the basis for further BPDUs it sends
  - the port where the BPDU with the smaller root ID was received is then marked as the root port
  - The remaining ports are placed in either blocked or forwarding states
- 3.4.1.4 Topology Changes
- topology change occurs when a port has entered the blocking or forwarding states
- when a bridge detects a connectivity change, the bridge notifies its parent bridges on the tree to the root by sending topology change notification BPDUs out of its root port
- Once informed of the topology change, the root bridge sets the TC bit field in subsequent periodic configuration messages
  - such messages are relayed by every bridge in the network and are received by ports in either the blocking or forwarding states
- The setting of this bit field allows bridges to reduce their aging time to that of the forward delay timer, on the order of seconds instead of 5 minutes normally recommended for the aging time
- 3.4.1.5 Example
- Linux bridge function disables STP by default
- brct1 stp br0 on
  - enables STP on the example bridge
  - can inspect the command using brct1 showstep br0
- STP setup for a simple bridge
  - br0 bridge device
    - includes bridge ID
    - derived from the smallest MAC address on the PC based bridge (port1)
    - The major configuration parameters are given in seconds
    - The flags values indicate a recent topology change
  - rest of the output describes per-port information eth0 (bridge port1) and eth1 (bridge port2)
- 3.4.1.6 Rapid Spanning Tree Protocol (RSTP)
- STP
  - change in topology is detected only by the failure to receive a BPDU in a certain amount of time
    - if the timeout is large, the convergence time could be larger than desired

- Rapid Spanning Tree Protocol (RSTP)
  - the main improvement in RSTP over STP is to monitor the status of each port and upon indication of failure to immediately trigger a topology change indication
  - RSTP uses all 6 bits in the Flag field of the BPDU format to support agreements between bridges that avoid some of the need for timers to initiate protocol operations
  - It reduces the normal STP five port states to three (discarding, learning, and forwarding)
    - discarding state in RSTP absorbs the disabled, blocking, and listening states in conventional STP
  - Also creates a new port role called an alternate port, which acts as an immediate backup should a root port cease to operate
  - uses only one type of BPDU
    - so there are no special topology change BPDUs
    - use version and type number 2 instead of 0
  - any switch detecting a topology change sends BPDUs indicating a topology change, and any switch detecting a topology change receiving them clears its filtering databases immediately
  - edge ports - those attached only to end stations and normal spanning tree ports on point-to-point links and shared links
  - In regular STP, BPDUs are ordinarily relayed by from a notifying or root bridge. In RSTP BPDUs are sent periodically by all bridges as “keepalives” to determine if connections to neighbors are operating properly.
  - if a bridge fails to receive an updated BPDU within three times the hello interval, the bridge conclude that it has lost its connection with its neighbor
  - topology changes are not induced as a result of edge ports being connected or disconnected as they are in regular STP
  - When a topology change is detected, the notifying bridge sends BPDUs with the TC bit field set, not only to the root but also to all other bridges. Doing so allows the entire network to be notified of the topology change must faster than with conventional STP. When a bridge receives these messages, it flushes all table entries except those associated with edge ports and restarts the learning process
  - RSTP has been extended to include VLANs a protocol called the Multiple Spanning Tree Protocol (MSTP). This protocol retains the RSTP BPDU format, so backward compatibility is possible but it also supports the formation of multiple spanning trees (one for each VLAN)
- 3.4.2 802.1ak: Multiple Registration Protocol (MRP)
  - provides a general method for registering attributes among stations in a bridged LAN environment
  - MVRP - for registering VLANs
    - once an end station is configured as a member of a VLAN, this information is communicated to its attached switch, which in turn propagates the fact of the station’s participation in the VLAN to other switches
    - this allows switches to augment their filtering tables based on station VLAN IDs and allows changes of VLAN topology without necessarily triggering a recalculation of the existing spanning tree via STP.
  - MMRP - for registering group MAC addresses
    - method for stations to register their interest in group MAC addresses
    - this information may be used by switches to establish the ports through which multicast traffic must be delivered
- 3.5 Wireless LANs - IEEE 802.11(Wi-Fi)

- wireless fidelity - 802.11
- Example figure
  - The IEEE 802.11 terminology for a wireless LAN. Access points (APs) can be connected using a distribution service (DS, a wireless or wired backbone) to form an extended WLAN (called an ESS). Stations include both APs and mobile devices communicating together that form a basic service set (BSS). Typically, an ESS has an assigned ESSID that functions as a name for the network.
- Stations are organized with a subset operating also as access points (AP)
- AP and its associated stations are called a basic service set (BSS)
- AP are generally connected to each other using a wired distribution service forming an extended service set (ESS)
  - commonly termed infrastructure mode
- 802.11 standard also provides for an ad hoc mode
  - in this configuration there is no AP or DS; instead, direct station-to-station (peer-to-peer) communication takes place
  - the STAs participating in an ad hoc network form an independent basic service set (IBSS)
- A WLAN formed from a collection of BSSs and/or IBSSs is called service set, identified by a service set identifier (SSID)
- Extended service set identifier (EESID) - is an SSID that names a collection of connected BSSs and is essentially a name for the LAN that can be up to 32 characters long

- 3.5.1 802.11 Frames
- one common overall frame format for 802.11 networks but multiple types of frames
- Example figure frame
  - 802.11 basic data frame format. The MPDU format resembles that of Ethernet but has additional fields depending on the type of DS being used among access points, whether the frame is headed to the DS or from it, and if frames are being aggregated. The QoS Control field is used for special performance features, and the HT Control field is used for control of 802.11n's "high throughput" features.
- The frame
  - The frame shown in the figure contains a preamble for synchronization which depends on the particular variant of 802.11 being used.
  - Physical Layer Convergence Procedure (PLCP) header provides information about the specific physical layer in a somewhat PHY-independent way.
    - generally transmitted at a lower data rate than the rest of the frame
    - Servers two purposes
      - to improve the probability of correct delivery (lower speeds tend to have better error resistance)
      - provide compatibility with and protection from interference from legacy equipment that may operate in the same area at slower rates
  - MAC PDU (MPDU) - corresponds to a frame similar to Ethernet, but with some additional fields
    - at the head of MPDU is the Frame Control Word - which includes a 2-bit type field identifying the frame type
    - Three types of frames: management frames, control frames, and data frames
      - each can have varying subtypes depending on the type
    - the contents of the remaining fields, if present, are determined by the frame type, which we discuss individually
- 3.5.1.1 Management Frames

    - Management frames are used for creating, maintaining, and ending associations between stations and access points.
    - Also used to determine whether encryption is being used
    - what the name of the network is
    - what transmission rates are supported
    - a common time base
    - these frames are used to provide information necessary when a Wi-Fi “scans” for nearby access points
- Scanning is the procedure by which a station discovers available networks and related configuration information
    - this involves switching to each available frequency and passively listening for traffic to identify available access points
- Stations may also actively probe for networks by transmitting a particular management frame (“probe request”) while scanning
- There are some limitations on such probe request to ensure that 802.11 traffic is not transmitted on a frequency that is being used for non-802.11 purposes
    - iwlist wlan0 scan
        - initiates a scan by hand on a Linux system
- scan
    - quality and signal level give indications of how well the scanning station is receiving a signal from the AP, although the meaning of these values varies among manufacturers
    - tsf (time sync function) value indicates the AP’s notion of time, which is used for synchronizing various features such as power saving mode
- When an AP broadcasts its SSID, any station may attempt to establish an association with the AP
    - when an association is established, most Wi-Fi networks today also set up the necessary configuration information to provide Internet access to the station
    - However, an AP’s operator may wish to control which stations make use of the network
        - some operators intentionally make this more difficult by having the AP not broadcast its SSID, as a security measure
- 3.5.1.2 Control Frames: RTS/CTS and ACKs
- Control frames are used to handle a form of flow control as well as acknowledgments for frames
    - flow control helps ensure that a receiver can slow down a sender that is too fast
    - acknowledgements help a sender know what frames have been received correctly
    - These concepts also apply to TCP at the transport layer
- 802.11 networks support optional request-to-send (RTS) / clear-to-send (CTS) moderation of transmission for flow control.
    - when these are enabled, prior to sending a data frame a station transmits an RTS frame, and when recipient is willing to receive additional traffic, it responds with a CTS
    - after the RTS/CTS exchange, the station has a window of time (identified in the CTS frame) to transmit data frames that are acknowledge when successfully received
- RTS/CTS
    - exchange helps to avoid the hidden terminal problem by instructing each station when it is permitted to transmit, so as to avoid simultaneous transmissions from stations that cannot hear each other
    - frames are short do not use the channel for long
    - AP generally initiates an RTS/CTS exchange for a packet if the size of the packet is large enough

- typically AP has a configuration option called the packet size threshold
- frames larger than the threshold cause an RTS to be sent prior to transmission of the data
- iwconfig wlan0 rts 250
  - iwconfig command can be used to set many variables, including the RTS and fragmentation thresholds
    - can be used to determine statistics such as the number of frame errors due to wrong network ID (ESSID) or wrong encryption key
    - it also gives the number of excessive retries (i.e., the number of retransmission attempts) - a rough indicator of the reliability of the link that is popular for guiding routing decisions in wireless networks
- In WLANs with limited coverage, where hidden terminal problems are unlikely to occur, it may be preferable to disable RTS/CTS by adjusting the stations RTS thresholds to be a high value.
  - This avoids the overhead imposed by requiring RTS/CTS exchanges for each packet
- In wired Ethernet networks, the absence of a collision indicates that a frame has been received correctly with high probability
- Wireless networks there is a wider range of reasons a frame may not be delivered correctly, such as insufficient signal or interference
- retransmission / acknowledgement (ACK) scheme
  - an acknowledgment is expected to be received within a certain amount of time for each unicast frame sent (802.11a/b/g) or each group of frames sent (802.11n or 802.11e with “block ACKs”). Multicast
  - multicast and broadcast frames do not have associated ACKs to avoid “ACK implosion”
  - Failure to receive an ACK within the specified time results in retransmission of the frame(s)
  - with retransmission, it is possible to have duplicate frames formed within the network
  - Retry bit - this field in the Frame Control Word is set when any frame represents a retransmission of a previously transmitted frame
    - a receiving station can use this to help eliminate duplicate frames
    - stations are expected to keep a small cache of entires indicating addresses and sequence/fragment numbers seen recently
      - when a received frame matches an entry, the frame is discarded
- the amount of time necessary to send a frame and receive an ACK for it relates to the distance of the link and the slot time
  - the time to wait for an ACK can be configured in most systems, although the method for doing so varies

- 3.5.1.3 Data Frames, Fragmentation, and Aggregation
- 802.11 supports frame fragmentation - which can divide frames into multiple fragments
  - also supports frame aggregation - which can be sed to send multiple frames together with less overhead
- fragmentation
  - each fragment has its own MAC header and trailing CRC and is handled independently of other fragments.
  - fragments to different destinations can be interleaved
  - can help improve performance when the channel has significant interference
  - applied only to frames with a unicast (non-broadcast or multicast) destination address
  - To enable this capability

    - Sequence Control field contains a fragment number (4 bits) and a sequence number (12 bits)
      - if a frame is fragmented, all frames contain a common sequence number value, and each adjacent fragment has a fragment number differing by 1
      - a total of 15 fragments for the same frame are possible, given the 4-bit-wide field.
      - The more Frag field in the Frame Control Word indicates that further fragments are yet to come
      - Terminal fragments have the bits set to 0
      - a destination defragments the original frame from fragments it receives by assembling the fragments in order based on fragment number order within the frame sequence number. Provided all fragments constituting a sequence number have been received and the last fragment has a More Frag field of 0, the frame is reconstructed and passed to higher-layer protocols for processing
  - often not used because requires tuning
    - or can worsen performance
  - The reason fragmentation can be useful is fairly simple exercise in probability
    - If the bit error rate (BER) is P, the probability of a bit being successfully delivered is (1-P) and the probability that N bits are successfully delivered is (1-P)^n. As n grows this value shrinks.
    - so if the BER is effectively 0, fragmentation only decreases performance by creating more frames to handle
- frame aggregation in two forms
  - aggregated MAC service data unit (A-MSDU) - allows for multiple complete 802.3 (Ethernet) frames to be aggregated within an 802.11 frame
    - For a single aggregate approach is technically more efficient
  - aggregated MAC protocol data unit (A-MPDU)- allows multiple MPDUs with the same source, destination, and QoS settings to be aggregated by being sent in short succession
    - different form of aggregation whereby multiple 802.11 frames, each with its own 802.11 MAC header and FCS and up to 4095 bytes are sent together
    - each constituent frame (subframe) carries its own FCS, it is possible to selectively retransmit only those subframes received with errors. This is made possible by the block acknowledgment facility in 802.11n
- 3.5.2 Power Save Mode and the Time Sync Function (TSF)
- 802.11 specification provides a way for stations to enter a limited power state, called power save mode (PSM)
  - When in PSM, an STA's outgoing frames have a bit set in the Frame Control Word
  - a cooperative AP noticing this bit being set buffers any frames for the station until the station requests them
  - APs ordinarily send out beacon frames (a type of management frame) indicating various things like SSID, channel, and authentication information
  - APs can also indicate the presence of buffered frames to a station by setting an indication in the Frame Control Word of the frames it sends
  - When stations enter PSM, they do so until the next AP beacon time, when they wake up and determine if there are pending frames stored at the AP for them
  - Time synchronization function (TSF)
    - each station maintains a 64-bit counter reference time that is synchronized with other stations in the network
    - any station that receives a TSF update - check to see if the provided value is larger than its own

    - automatic power save delivery (APSD)
        - useful for small power-constrained devices, as they need not necessarily awaken at each beacon interval as they do in conventional 802.11
    - spatial multiplexing power save mode - allows a station equipped with multiple radio circuits operating together to power down all but one of the circuits until a frame is ready
    - Power Save Multi-Poll (PSMP) - that provides a way to schedule transmissions of frames in both directions at the same time
- 3.5.3 802.11 Media Access Control
- In essence, the medium is effectively simplex, and multiple simultaneous transmitters must be avoided, by coordinating transmissions in either a centralized or a distributed manner
- three approaches to control sharing of the wireless medium
    - point coordination function (PCF)
    - distributed coordinating function (DCF)
        - form of CSMA/CA for contention-based access to the medium used for both infrastructure and ad hoc operation
        - With CSMA/CA stations listen to see if the medium is free and, if so, may have an opportunity to transmit. If not, they avoid sending for a random amount of time before checking again to see if the medium is free
        - distributed inter-frame space (DIFS) - time that stations wait when ready to send to allow higher priority stations to access the channel
            - If the channel becomes busy during the DIFS period, a station starts the waiting period again
            - When the medium appears idle, a would be transmitter initiates the collision avoidance/ backoff procedure
            - Procedure is initiated after a successful transmission is indicated by the receipt of an ACK
            - In the case of an unsuccessful transmission, the back procedure is initiated with a different timing
                - Extended interframe space (EIFS)
    - hybrid coordination function (HCF)
- 3.5.3.1 Virtual Carrier Sense, RTS/CTS, and the Network Allocation Vector (NAV)
- 802.11 MAC protocol
    - virtual carrier sense mechanism operates by observing the Duration field present in each MAC frame
    - duration field - is present in both RTS and CTS frames optionally exchanged prior to transmission, as well as conventional data frames, and provides an estimate of how long the medium will be busy carrying the frame
- Network Allocation Vector (NAV) - local counter kept by each station
    - estimates how long the medium will be busy carrying the current frame, and consequently how long it will need to wait before attempting its next transmission
- 3.5.3.2 Physical Carrier Sense (CCA)
    - each PHY specification is required to provide a function for assessing whether the channel is clear based upon energy and waveform recognition
    - clear channel assessment (CCA)
        - implementation PHY-dependent
        - represents the physical carrier sense capability for the 802.11 MAC to understand whether the medium is currently busy
        - used in conjunction with the NAV to determine when a station must defer (wait) prior to transmission

- 3.5.3.3 DCF Collision Avoidance/Backoff Procedure
  - a station defers access prior to transmission
  - because many stations may have been waiting for the channel to become free, each station computes and waits for a backoff time prior to sending.
  - backoff time is equal to the product of a random number and the slot time
    - slot time - PHY-dependent but is generally a few tens of microseconds
  - collision detection is not practical because it is difficult for a transmitter and receiver to operate simultaneously in the same piece of equipment and hear any transmissions other than its own, so collision avoidance is used instead.
- 3.5.3.4 HCF and 802.11e/n QoS
  - hybrid coordination function - supports both contention-based and controlled channel access
    - HFCA-controlled channel access (HCCA)
    - enhanced DCF channel access (EDCA)
      - builds upon the basic DCF access
      - eight user priorities that are mapped to four access categories (ACs)
        - labeled 1 through 7 being the highest priority
        - four priorities intended for the background, best effort, video, and audio traffic
        - priorities 1 and 2
          - intended for the background AC
        - priorities 0 and 3
          - intended for the best effort AC
        - priorities 4 and 5
          - are for the video
        - priorities 6 and 7
          - are intended for the voice AC
      - for each AC, a variant of a DCF contends for channel access credits called transmit opportunities (TXOPs), using alternative MAC parameters that tend to favor the higher-priority traffic
    - admission control - which may deny connectivity entirely under high load
  - hybrid coordinator (HC) - located within an AP and has priority to allocate channel accesses
  - traffic specification - prior to transmission issued by a station for its traffic and use UP values between 8 and 15. The HC can allocate reserved TXOPs to such request to be used during short -duration controlled access phases of frame exchange that take place before EDCA-based frame transmission
  - HC can also deny TXOPs to TSPECs based on admission control policies set by the network administrator
  - Note that a single network comprising QSTAs and conventional stations can have both HCF and DCF running simultaneously by alternating between the two, but ad hoc networks do not support the HC and thus do not handle TSPECs and do not perform admission control.
- 3.5.4 Physical-Layer Details: Rates, Channels, and Frequencies
- Table parts of the 802.11 standard that describe the physical layer
- 3.5.4.1 Channels and Frequencies
- regulatory bodies divide the electromagnetic spectrum into frequency ranges allocated for various uses across the world
  - for each range and use a license may or may not be required, depending on local policy
- For typical Wi-Fi networks, an AP has its operating channel assigned during installation, and client stations change channels in order to associate with the AP.

- When operating in ad hoc mode, there is no controlling AP, so a station is typically hand-configured with the operating channel
- 3.5.4.2 802.11 Higher Throughput/802.11n
- 802.11n support higher throughput, it incorporates support for multiple input, multiple output management of multiple simultaneously operating data streams carried on multiple antennas, called spatial streams
  - up to four spatial streams are supported on a given channel
- 802.11n also improves single-stream performance by using a more efficient modulation scheme
  - uses MIMO- orthogonal frequency division multiplexing (OFDM)
  - also reduces the guard interval (GI, a forced idle time between symbols)
  - Some 77 combinations of modulation and coding options are supported by 802.11n
    - including 8 options for a single stream
    - 24 using the same or equal modulation (EQM) on all streams
    - 43 using unequal modulation (UEQM) on multiple streams
  - the more high-performance and complex a modulation scheme, the more vulnerable it tends to be to noise and interference
  - Forward error correction (FEC) - includes a set of methods whereby redundant bits are introduced at the sender that can be used to detect and repair bit errors introduced during delivery.
    - code rate - is the ratio of the effective useful data rate to be the rate imposed on the underlying communication channel.
- 802.11n may operate in 3 modes
  - greenfield mode - known only to 802.11n equipment and does not interoperate with legacy equipment
  - non-HT mode - disables all 802.11n features but remains compatible with legacy equipment
  - HT-mixed mode - supports both 802.11n and legacy operation, depending on which stations are communicating
- Optional BSS feature called phased coexistence operation (PCO) allows an AP to periodically switch between 20MHz and 40MHz channel widths, which can provide better coexistence between 802.11n APs operating near legacy equipment at the cost of some additional throughput
- 3.5.5 Wi-Fi Security
- wired equivalent privacy (WEP) - shown to be so weak that some replacement was required
- Wi-Fi protected access (WPA) - replaced the way keys are used with encrypted blocks
  - Temporal Key Integrity Protocol (TKIP) - ensures, among other things, that each frame is encrypted with a different encryption key.
- encryption techniques are aimed at providing privacy between the station and the AP
- 3.5.6 Wi-Fi Mesh (802.11s)
- mesh - operation, wireless stations can act as data-forwarding agents (like APs)
- mesh stations - are a type of QoS STA and may participate in HWRP (Hybrid Wireless Routing Protocol) or other routing protocols, but compliant nodes must include an implementation of HWRP and the associated airtime link metric.
- mesh nodes - coordinate using EDCA or may use an optional coordinating function called mesh deterministic access
- mesh points - are those nodes that form mesh links with neighbors
  - those that also include AP functionality are called mesh APs (MAPs)
  - conventional 802.11 stations can use either APs or MAPs to access the rest of the wireless LAN

- Simultaneous Authentication of Equals (SAE) - stations are treated as equals , and any station that first recognizes another may initiate a security exchange (or this may happen simultaneously as two stations initiate an association)
- 3.6 Point-to-Point Protocol (PPP)
- PPP - is a popular method for carrying IP datagrams over serial links — from low-speed dial-up modems to high-speed optical links
- Link Control Protocol (LCP) - basic method to establish a link
- 3.6.1 LCP
- This portion of PPP is used to establish and maintain a low-level two-party communication path over a point-to-point link
- point-to-point link must support bidirectional operation
- Typically LCP establishes a link using HDLC (High-Level Data Link Control)
- Example figure
  - PPP basic frame format was borrowed from HDLC. It provides a protocol identifier, payload area, and 2- or 4-byte FCS. Other fields may or may not be present, depending on compression options
- asynchronous links PPP uses character stuffing - flag character and escape character are replaced if appear in the frame. 0x7E and 0x7D
- synchronous links PPP uses bit stuffing - arranges for a 0 bit to be inserted after any contiguous string of five 1 bits appearing in a place other than the flag character itself
- Format PPP
  - Flag field
  - Address
    - specifies which station is being addressed
    - PPP is only concerned with a single destination, this field is always defined to have the value 0xFF
  - Control fields
    - used to indicate frame sequencing and retransmission behavior
    - fixed value 0x03
  - Address and Control Field Compression (ACFC) - address and control fields omitted during transmission. Which essentially eliminates the two fields.
  - Protocol - indicates the type of data being carried
    - network layer protocols - range 0x0000-0x3FFF
    - data belong to associated NCP - range 0x8000-0xBFFF
    - “low volume” protocols - range 0x4000-0x7FFF
    - control protocols - range 0xC000-0XEFFF
    - Protocol Field Compression (PFC) - protocol field can be compressed to a single byte
      - option negotiated successfully during link establishment
      - range 0x0000-0x00FF
        - includes most of the popular network layer protocols
        - LCP packets always use the 2-byte uncompressed format
  - 16-bit FCS - covering the entire frame except the FCS field and Flag bytes
- 3.6.1.1 LCP Operation
- LCP Packet
  - Flag 0x7E
  - Addr 0xFF
  - Control 0x03
  - Protocol 0xC021
  - Code

    - gives the type of operation being either requested or responded to
    - Generally
      - ACK messages indicate acceptance of a set of options
      - NACK messages indicate a partial rejection with suggested alternatives
      - REJECT rejects one or more options entirely
  - Indent
    - is a sequence number provided by the sender of LCP request frames and is incremented for each subsequent message
  - Length
    - gives the length of the LCP packet in bytes and is not permitted to exceed the link's maximum received unit (MRU)
  - LCP Data
    - to bring up a point-to-point link to minimal level
    - Configure messages - cause each end of the link to start the basic configuration procedure and establish agreed-upon options
    - Termination messages - are used to clear a link when complete
    - Request/Reply - messages may be exchanged anytime a link is active by LCP in order to verify operation of the peer
    - Discard Request - message can be used for performance measurement; it instructs the peer to discard the packet without responding
    - Identification and Time-Remaining - messages are used for administrative purposes: to know the type of the peer system and to indicate the amount of time allowed for the link to remain established.
  - Pad
  - FCS
  - Flag
- 3.6.1.2 LCP Options
- several options can be negotiated by LCP as it establishes a link for use by one or more NCPs
  - Asynchronous Control Character Map (ACCM) - "asyncmap" option defines which control characters need to be "escaped" as PPP operates
    - escaping a character means that the true value of the character is not sent, but instead the PPP escape character (0x7D) is stuffed in front of a value formed by XORing the original control character with the value 0x20
  - Link Quality Reports (LQRs) - peer is asked to provide LQRs at some periodic rate
    - include the following information
      - magic number
      - the number of packets and bytes sent and received
      - the number of incoming packets with errors and the number of discarded packets
      - total number of LQRs exchanged
    - can terminate link if quality fails to meet some threshold
  - callback capability
    - a PPP dial-up callback client calls in to a PPP callback server, authentication information is provided, and the server disconnects and calls the client back
      - useful in situations where call toll charges are asymmetric or for some level of security
- block size - certain minimum number of bytes
- padding
  - padding is included beyond the data area and prior to the PPP FCS field
- self-describing padding - alters the value of padding to be nonzero

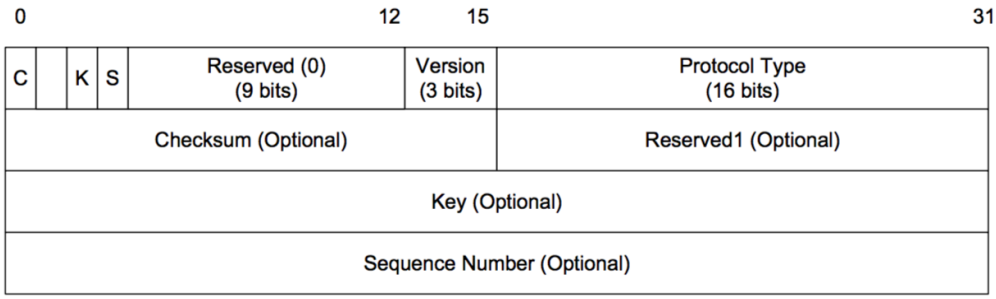
    - each byte gets the value of its offset in the pad area
- maximum pad value (MPV)
    - largest pad value allowed for this association
- 3.6.2 Multilink PPP (MP)
- special option to PPP can be used to aggregate multiple point-to-point links to act as one
- bundle - an aggregated link operates as a complete virtual link an can contain its own configuration information
    - comprised of a number of member links - each member link may also have its own set of options
- bank teller's algorithm - simply alternates packets across the member links
    - may lead to reordering of packets, which can have undesirable performance impacts on other protocols
- MP method
    - places a 2- or 4-byte sequencing header in each packet, and the remote MP receiver is tasked with reconstructing the proper order
- MP
    - requested by including an LCP option called the multilink maximum received reconstructed unit (MRRU)
        - can act as a sort of larger MRU applying to the bundle
        - Member links in the same bundle are identified by the LCP endpoint discriminator option
    - basic method of establishing member links expects member links are going to be used symmetrically about the same number of fragments will be allocated to each of a fixed number of links
    - More sophisticated allocations
        - Bandwidth Allocation Protocol (BAP) - can be used to dynamically add or remove links from a bundle
            - includes three packet types: request, response, and indication
                - requests - add a link to a bundle
                - indications - convey the results of attempted additions back to the original requester and are acknowledged
                - responses - either ACKs or NACKs for these requests
        - Bandwidth Allocation Control Protocol (BACP) - can be used to exchange information regarding how links should be added or removed using BAP
- 3.6.3 Compression Control Protocol (CCP)
- CCP
    - can be negotiated only once the link has entered the Network state
    - behaves like NCP
    - uses the same packet exchange procedures and formats as LCP except the Protocol field is set to 0x80FD
    - some special options in addition to the common Code field values two new operations are defined: reset-request(0x0e) and reset-ACK(0x0f)
    - One or more compressed packets may be carried within the information portion of a PPP frame
    - Compressed frames carry the protocol field value of 0x00FD, but the mechanism used to indicate the presence of multiple compressed datagrams is dependent on the particular compression algorithm used.
    - If used in conjunction with MP and only on member links, the protocol field is set to 0x00FB
- 3.6.4 PPP Authentication
- basic PPP specification has a default of no authentication

- Password authentication protocol - simplest and least secure scheme
  - one peer requests the other to send a password, and the password is so provided
  - password is sent unencrypted over the PPP link, any eavesdropper on the line can simply capture the password and use it later
  - PAP packets are encoded as LCP packets with the Protocol field value set to 0xC023
- Challenge-Handshake Authentication Protocol -
  - a random value is sent from one peer (called the authenticator) to the other.
  - a response is formed by using a special one-way (i.e., not easily invertible) function to combine the random value with a shared secret key (usually derived from a password) to produce a number that is sent in response
  - upon receiving the response, the authenticator can determine with a very high degree of confidence that is peer possesses the correct secret key.
  - The protocol never sends the key or password over the link in a clear (unencrypted) form, so any eavesdropper is unable to learn the secret.
  - because a different random value is used each time, the result of the function changes for each challenge/response, so the values an eavesdropper may be able to capture cannot be reused (played back) to impersonate the peer.
  - However, vulnerable to a “man in the middle attack”
- EAP authentication framework available for many different network types
  - defines a message format for carrying a variety of specific types of authentication formats, but additional specifications are needed to define how EAP messages are carried over particular types of links
  - when used with PPP the authentication operation may be postponed until Auth state (just before the network state)
- 3.6.5 Network Control Protocols (NCPs)
- IP Control Protocol - NCP can be used on a PPP link
- Once LCP has completed its link establishment and authentication, each end of the link is in the Network state and may proceed to negotiate a network-layer association using zero or more NCPs
- IPCP
  - the standard NCP for IPv4, can be used to establish IPv4 connectivity over a link and configure Van Jacobson header compression
  - IPCP packets may be exchanged after the PPP state machine has reached the network state
  - Packets same as LCP except protocol field is set to 0x8021
  - Code field is limited to the range 0 - 7
  - Can negotiate a number of options: IP compression protocol, IPv4 address, Mobile IPv4 and other options available for learning the location of primary and secondary domain name servers
- IPv6CP
  - uses the same packet exchange and format as LCP, except it has two different options: interface identifier and IPv6-compression-protocol
- 3.6.6 Header Compression
- it is useful to have a way of compressing the headers of these higher-layer protocols (or eliminating them) so that fewer bytes need to be carried over relatively slow point-to-point links
- VJ compression
  - portions of a higher-layer headers are replaced with small,1-byte connection identifier

    - when the non changing values are sent over a link once (or a small number of times) and kept in a table, a small index can be used as a replacement for the constants in subsequent packets.
- IP header compression
    - provides a way to compress the headers of multiple packets using both TCP or UDP transport-layer protocols and either IPv4 or IPv6 network-layer protocols
    - points out the necessity of some strong error detection mechanism in the underlying link layer because erroneous packets can be constructed at the egress of a link if compressed header values are damaged in transit.
- Robust Header Compression
    - further generalizes IP header compression to cover more transport protocols and allows more than one form of head compression to operate simultaneously. Like the IP header compression mentioned previously, it can be used over various types of links, including PPP
- 3.6.7 Example
- Debugging output of a PPP server interacting with a client over a dial-in modem
    - PPP server process creates a (virtual) network interface called ppp0 which is awaiting an incoming connection on the dial-up modem attached to serial port ttyS0
    - Once the incoming connection arrives, the server request an asyncmap of 0x0, EAP authentication, PFC, and ACFC
    - Client refuses EAP authentication and instead suggests MS-CHAP-v2
    - Server then tries again, this time using MS-CHAP-v2, which is then accepted and acknowledged
    - Incoming request includes CBCP; an MRRU of 1614 bytes, which is associated with MP support; and endpoint ID
    - The server rejects the request for CBCP and multilink operation
    - The endpoint discriminator is once again sent by the client , this time without the MRRU, and is accepted and acknowledged
    - Server sends CHAP challenge with the name dialer
    - Before response to challenge arrives, two incoming identity messages arrive, indicating that the peer is identified by the strings
    - Finally, the CHAP response arrives and is validated as correct, and an acknowledgement indicates that access is granted. PPP then moves on the Network state
    - Once in Network state, the CCP, IPCP, and IPV6CP NCPs are exchanged
    - CCP attempts to negotiate Microsoft Point-to-Point Encryption
    - MPPE is somewhat of an anomaly, as it is really an encryption protocol, and rather than compressing the packet it actually expands it by 4 bytes
    - Note during the middle of negotiation, the client attempts to send an IPCP request, but the server responds with an unsolicited TermAck (a message defined within LCP that ICPC adopts)
        - termACk used to indicate to the peer that the server is “in need of renegotiation”
    - After successful negotiation of MPPe, the server request the use of VJ header compression and provides it IPv4 and IPv6
    - IPv4 initially rejected
    - IPv6 accepted and acknowledged
    - Client ACKs both IPv4 and IPv6 addresses of the server
    - Client again requests IPv4 addresses of 0.0.0.0, which is rejected in favor of 192.168.0.1
- 3.7 Loopback
- summary

  - many cases clients may wish to communicate with servers on the same computer using Internet protocols such as TCP/IP
  - most implementations support a network-layer loopback capability that typically takes the form of a virtual loopback network interface
  - It acts like a real network interface but is really a special piece of software provided by the operating system to enable TCP/IP and other communications on the same host computer.
- IPv4
  - addresses starting with 127 are reserved for this
  - UNIX like systems including Linux assign the IPv4 address of 127.0.0.1 to the loopback interface and assign it the name localhost
- Linux
  - loopback interface is called lo
  - ifconfig lo
- 3.8 MTU and Path MTU
- Maximum transmission unit - fixed upper limit on the size of the frame available for carrying PDUs
- If IP has a datagram to send, and the datagram is larger than the link layer's MTU, IP performs fragmentation, breaking the datagram up into smaller pieces (fragments), so that each fragment is smaller than the MTU
- minimum MTU across the network path comprising all the links is called the path MTU
- path MTU discovery (PMTUD) - used to determine the path MTU at a point in time
- 3.9 Tunneling Basics
- establish a virtual link between one computer and another across the Internet or other network
- Tunneling - carrying low layer traffic in high layer packets
- Tunneling allows for the formation of overlay networks
  - great variety of methods for tunneling packets of one protocol and/or layer over another

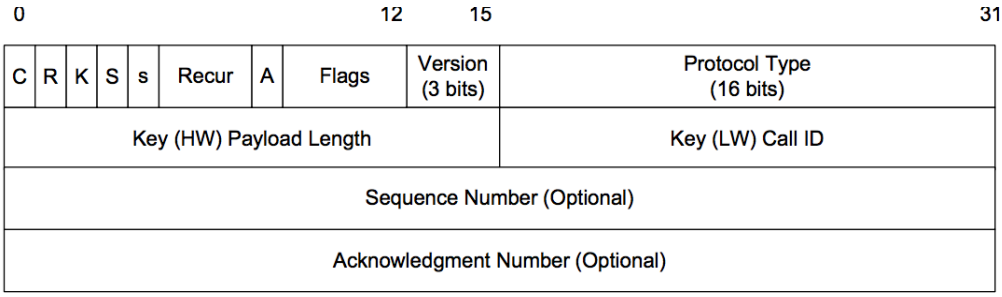
RFC 2890



- Generic Routing Encapsulation (GRE)
  - typically used within the network infrastructure to carry traffic between ISPs or within an enterprise intranet to serve branch offices and are not necessarily encrypted, although GRE tunnels can be combined with IPsec.
  - GRE header
    - C indicates whether a checksum is present
      - If the checksum field is present, the Reserved1 field is also present and is set to 0
    - Key and sequence fields are present if K and S fields are set to 1
      - if present, the key field is arranged to be a common value in multiple packets, indicating that they belong to the same flow of packets

- Sequence number field is used in order to reorder packets if they should become out of sequence

RFC 2637



- PPTP (Point-to-Point Tunneling Protocol)
  - most often used between users and their ISPs or corporate intranet and is encrypted
  - essentially combines GRE with PPP, so GRE can provide the virtual point-to-point link upon which PPP operates
  - often used to carry layer 2 frames so as to emulate a direct LAN (link-layer) connection
  - PPTP header
    - includes an extra R, s, and A bit fields, additional Flags field, and Recur field
    - K,S, and A fields indicate that the Key, Sequence Number, and Acknowledgement Number fields are present
    - the value of the Sequence Number field holds the largest packet number seen by the peer
- 3.9.1 Unidirectional Links
- unidirectional links (UDLs) - link to be used operates in only one direction
- many of the protocols describe so far do not operate properly in such circumstances because they require exchanges of information
- to deal with this situation, a standard has been created whereby tunneling over a second Internet can be combined with operation of the UDL
- Satellite example
  - Internet connection that uses a satellite of downstream traffic and a dial-up modem link for upstream traffic
  - To establish and maintain tunnels automatically at the receiver
  - Dynamic Tunnel Configuration Protocol (DTCP)
    - involves sending multicast Hello messages on the downlink so that any interested receiver can learn about the existence of the UDL and its MAC and IP addresses.
    - in addition Hello messages indicate a list of tunnel endpoints within the network that can be reached by the user's secondary interface
    - After the user selects which tunnel endpoint to use, DTCP arranges for return traffic to be encapsulated with the same MAC type as the UDL in GRE tunnels
    - The service provider arranges to receive these GRE-encapsulated layer 2 frames, extract them from the tunnel, and forward them appropriately
    - Although the upstream side of the UDLs requires manual tunnel configuration, the downstream side, which includes many more users, has automatically configured tunnels
- Automatic tunnel configuration techniques
  - 6to4
    - IPv6 packets are tunneled over an IPv4 network using the encapsulation specified in ….
  - Teredo
    - IPv6 transition using automatically configured tunnels

- Teredo tunnels IPv6 packets over UDP/IPv4 packets

- 3.10 Attacks on the Link Layer
  - wired Ethernet
    - interfaces can be placed in promiscuous mode, which allows them to receive traffic even if it is not destined for them
    - attack on switches
      - switches hold tables of stations on a per-port basis. If these tables are able to be filled quickly, it is conceivable that the switch might be forced into discarding legitimate entries, leading to service interruption for legitimate stations
      - related but worse attack can be mounted using the STP.
        - attacking a station can masquerade as a switch with a low-cost path to the root bridge and cause traffic to be directed toward it
  - Wi-Fi networks
    - more sophisticated set of attacks on Wi-Fi involves attacking the cryptographic protection, especially the WEP encryption used on many early access points
  - PPP Links
    - if the attacker can gain access to the channel between the two peers
    - for simple authentication mechanisms sniffing can be used to capture the password in order to facilitate illegitimate subsequent use
    - depending on the type of higher-layer traffic being carried over the PPP link, additional unwanted behaviors can be induced
  - Tunneling
    - can play the role of both target and tool
    - target
      - tunnels pass through a network (often the Internet) and thus are subject to being intercepted and analyzed
      - configured tunnel endpoints can also be attacked, either by attempting to establish more tunnels than the endpoint can support (a DoS attack) or by attacking the configuration itself
        - if the configuration is compromised, it may be possible to open an unauthorized tunnel to an endpoint.
        - at this point the tunnel becomes a tool rather than a target, and protocols such as L2TP can provide a convenient protocol-independent method of gaining access to private internal networks at the link layer.
        - Example
          - GRE-related attack, for example, traffic is simply inserted in a non encrypted tunnel, where it appears at the tunnel endpoint and is injected to the attached “private” network as though it were sent locally

- 3.11 Summary
  - Link layer - the lowest layer in the Internet protocol suite
  - VLANs, priorities, link aggregation, and frame formats
  - Point-to-point links and PPP protocol
  - loopback interface - loopback data has been completely processed by the transport layer and by the IP when it loops around to go up to the protocol stack
  - MTU and concept of a path MTU
  - Tunneling - involves carrying lower-layer protocols in higher-layer packets
    - allows for the formation of overlay networks, using tunnels over the Internet as links in another level of network infrastructure
  - Link layer - as target or tool

- One reason for the success of TCP/IP is its ability to work on top of almost any link technology
- IP requires only that there exists some path between sender and receiver across a cascade of intermediate links