• IP Internetworking

- an internetwork (internet) is a collection of networks interconnected to provide some host-to-host pkt delivery service
  (network refers to a directly connected or switched network using a single technology eg Ethernet, ATM; also called a physical network)
- nodes that interconnect networks are called routers
- the Internet Protocol (IP) is the main means at present for building scalable & heterogeneous internetworks
- IP service model
  - the service model for an internetwork must be compatible with that of underlying networks
  - best effort (unreliable) datagram (connectionless) model of data delivery (simple service, easy for underlying networks to provide → IP can "run over anything")
  - IP datagram consists of a header followed by a number of data bytes
  - the header fields (IPv4) include
    - version (current version of IP is 4, some routers support version 6)

- header length
- type of service
- length of datagram
- fragmentation info
- time to live (TTL), in hops
- protocol (demux key identifying higher-level protocol to which the IP pkt should be passed)
- checksum (for IP header)
- source & destination addresses (so that pkt can be delivered & destination can reply)

- IP service model supports <u>fragmentation & reassembly</u> for networks with different MTUs
  - each fragment is a self-contained IP datagram transmitted independently of other fragments
  - reassembly is done at the receiving host (if a fragment is lost, the receiver tries unsuccessfully to reassemble the datagram; it eventually gives up & reclaims the memory used)
  - hosts are encouraged to perform path MTU discovery to discover the smallest MTU in the source-receiver path & avoid fragmentation

- IP addressing (IPv4)
  - globally unique address for each host/router interface
  - hierarchical address structure
    - network part: identifies the network to which a host (or router interface) is attached
    - host part: identifies each host uniquely on the network
    - can have additional levels of hierarchy
  - address is 32 bits (4 bytes) long, usu. written with each byte in decimal, separated by dots, eg 192.32.34.139
  - original 'classful' addressing
    - if the most significant bit is 0, it is a class A address with 7 bits for the network part & 24 bits for the host part (intended for WANs)
    - if the most significant 2 bits are 10, it is a class B address with 14 bits for the network part & 16 bits for the host part (intended for site/campus-sized networks)
    - if the most significant 3 bits are 110, it is a class C address with 21 bits for the network part & 8 bits for the host part (intended for LANs)
    - address assignment inefficiency caused by insufficient flexibility (only 3 very different network address sizes)

- current Internet address assignment uses Classless Interdomain Routing (CIDR)
  - balances efficiency of address assignment with routing complexity
  - network numbers (called prefixes) can be of any length
  - the x most significant bits of an address of the form a.b.c.d/x constitute the network prefix
  - an organization typically has a range of addresses with a common prefix; the remaining 32-x bits may have additional hierarchical structure specifying subnets within the organization
  - eg. an ISP with the address block 200.23.16.0/20 might allocate portions of its address block to customer organizations as follows:

	addr block	prefix
ISP	200.23.16.0/20	11001000 00010111 0001
org 1	200.23.16.0/23	11001000 00010111 0001000
org 2	200.23.18.0/23	11001000 00010111 0001001

  - the use of a single prefix to specify routes to multiple networks (address aggregation / route aggregation / route summarization) improves scalability

- IP Datagram forwarding

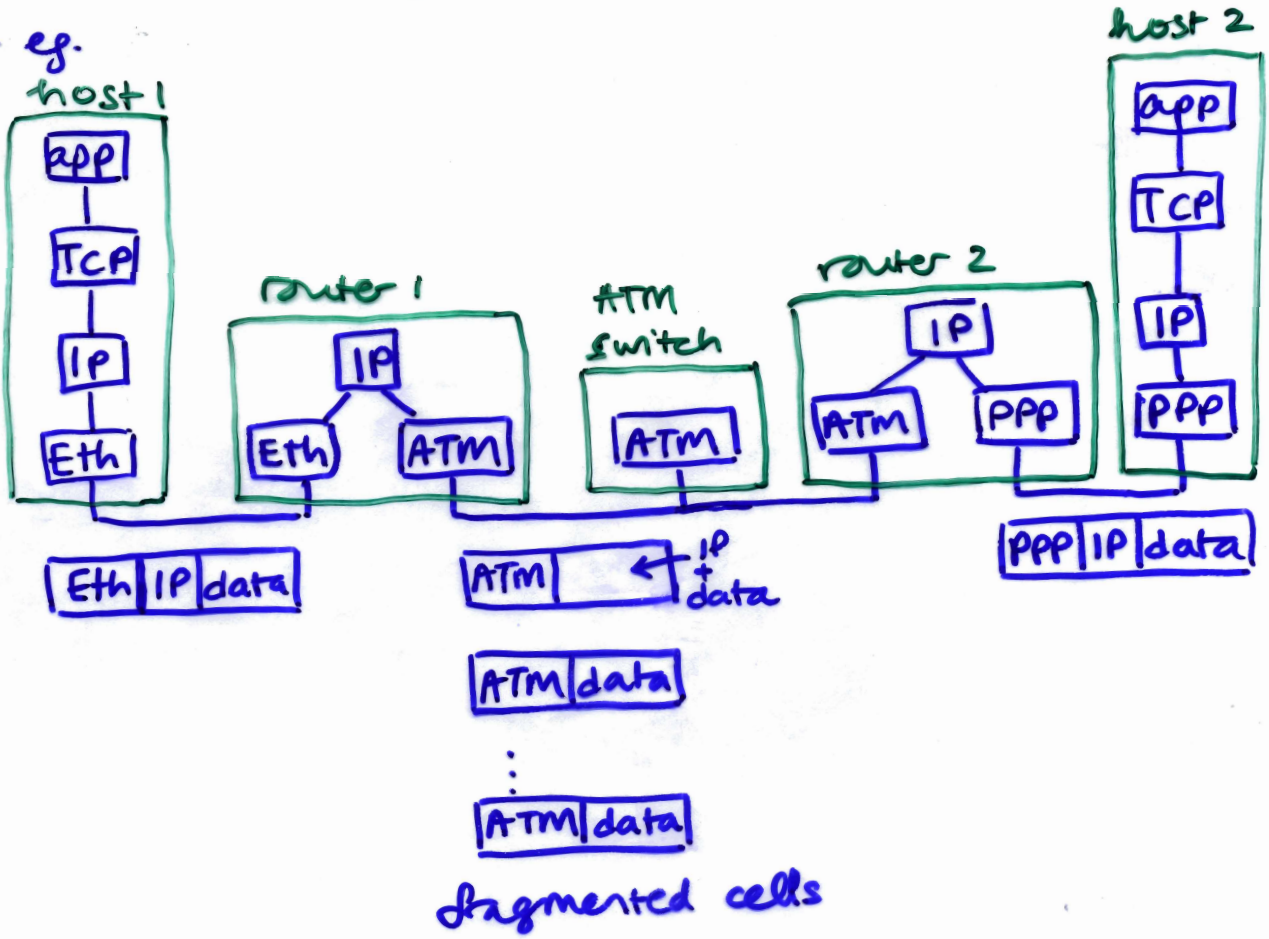
- all hosts & routers that have the same network part of their address are connected to the same physical network & can communicate by sending link-layer frames over this network
- as an IP datagram is sent from a source to a destination host, the source & intermediate routers (if any) first check whether the destination is connected to the same physical network as itself
  - this is done by comparing the network part of the destination address with the network part of the address of each of its own interfaces
- if there is a match, the IP address of the destination is translated to a link-level address for the physical network, & the IP datagram is encapsulated in a frame with the link-level address & sent over the physical network
  - for LANs with broadcast support, eg Ethernet & token ring, the Address Resolution Protocol (ARP) is used to enable each host on the network to build up a table of (IP addr, link-level addr) mappings (which are periodically timed out)

- a node learns the link-level address for a target IP address by broadcasting an ARP query on the network
- the host whose IP address matches the query sends a response message containing its link-layer address back to the originator of the query, which adds the info to its ARP table
- the query message also contains the sender's IP & link-level addresses, which are stored by the target host (& other hosts, if they have an existing entry for the sender, will refresh it)
- for ATM networks which do not support broadcast, an ARP server is used (nodes register their IP & ATM addresses with the ARP server when they boot)

• if the source / router finds that the destination is not on the same physical network as itself, it forwards the datagram to a next hop router
- a router chooses the next hop IP router by looking for the largest matching prefix in its forwarding table
- typically there is a default router that is used if none of the table entries match

- a host may have a default router & nothing else
- the IP datagram is reencapsulated for every physical network over which it travels eg. for an ATM network, ATM ARP is used to find the ATM address of the next hop IP router (the exit router), the appropriate VCI is determined / set up, & the segmented cells are forwarded at the ATM layer; at the exit router, the cells are reassembled, & the IP datagram is extracted & passed up to the IP layer

- eg.



- Error reporting

- IP has a companion protocol, the Internet Control Message Protocol (ICMP) that defines

a) a set of error messages sent back to the source host when a router / host cannot process an IP datagram successfully
   - eg. destination host unreachable, reassembly failed, TTL reached 0, IP checksum failed

b) a set of control messages that a router can send to a host
   - eg. ICMP- redirect tells a source that another router has a better route to a particular destination than the router (itself) being used