Design and Engineering of Computer Systems

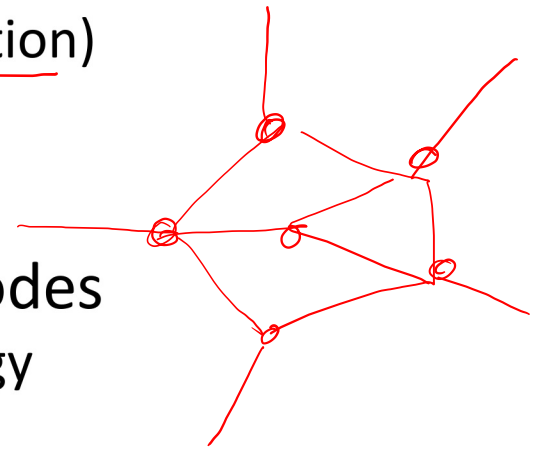
Lecture 22: Internet Routing and Forwarding

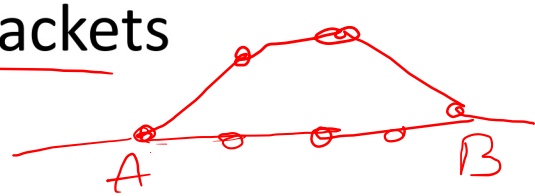
Mythili Vutukuru

IIT Bombay

IP layer / network layer / L3

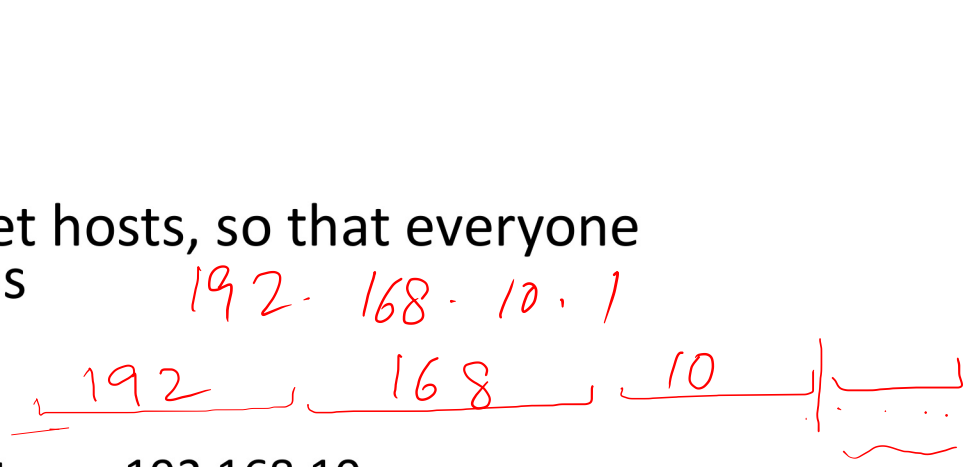
- Every network interface connected to the Internet has an IP address
  - End host with multiple interfaces will have multiple IP addresses
  - 32-bit IPv4 address, e.g., written as 192.168.10.1 (dotted decimal notation)
  - Or, 128-bit IPv6 address (introduced because of shortage of IPv4)
- End hosts connected to each other via series of IP routers
- IP routers run routing protocols that discover paths between nodes
  - Distributed/decentralized protocols, no one is told full network topology
  - Builds routing table = list of all known routes to every destination
- Among many known routes, best one is chosen for forwarding packets
  - Forwarding table = best path to a destination, next IP hop to go to
- Sometimes, IP routers may not be directly connected by a link
  - Link layer provides abstraction of direct connection between nodes at IP layer





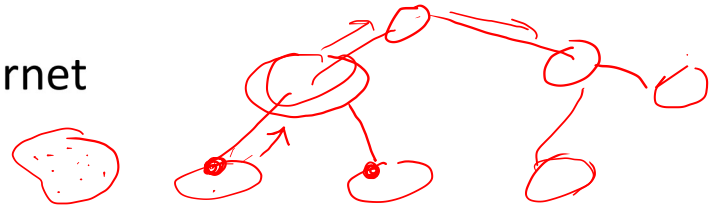
IP address and prefix

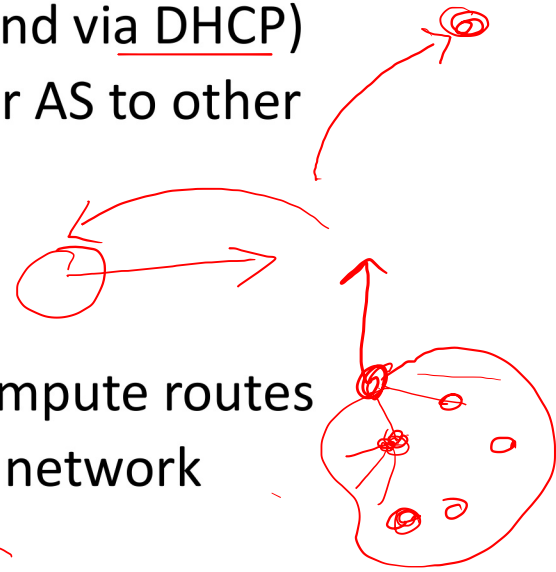
- Routing protocols exchange information about Internet hosts, so that everyone can learn about the network topology, compute routes
  - At what granularity do we exchange information?
- IP addresses are grouped into IP prefixes or subnets
  - 192.168.10.0/24 denotes 256 IP addresses whose first 24 bits are 192.168.10
  - 192.0.0.0/8 contains 256 "/16" prefixes
  - Subnet mask 255.0.0.0 denotes a prefix of length 8
- IP addresses are assigned to organizations at granularity of IP prefixes
  - Organization can further split a prefix into smaller prefixes for smaller subnets
- Routing protocols exchange information at the granularity of IP prefixes
- When a router receives IP datagram, it finds the matching prefix containing the destination IP address, uses route corresponding to this prefix
  - What if multiple IP prefixes match? Pick the longest (most specific) prefix



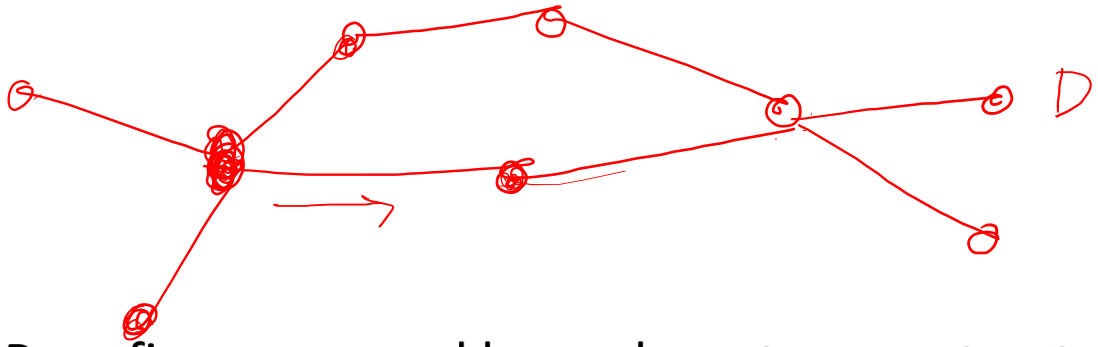
Internet topology

- The Internet is composed of multiple smaller independent networks called autonomous systems (AS)
  - Can be organizations with end users (clients and servers) or organizations that connect end users (Internet Service Providers or ISPs)
  - Multiple tiers of ISPs connect various "stub" organizations to form the Internet
- Each AS has one or more IP prefixes assigned to it
  - By whom? Internet registries allocate IP addresses to organizations
- Within AS, distribute addresses to hosts statically or dynamically (on-demand via DHCP)
- Every AS has one or more border routers that advertise the prefixes of their AS to other neighboring border routers
  - Stubs in the Internet announce routes to their ISPs (for a payment)
  - ISPs exchange information amongst each other, compute network-wide routes
  - Traffic flows in opposite direction of route announcements
- Within AS, internal IP routers exchange information on smaller prefixes, compute routes
- Hierarchical routing: packet first reaches border router, then routed within network



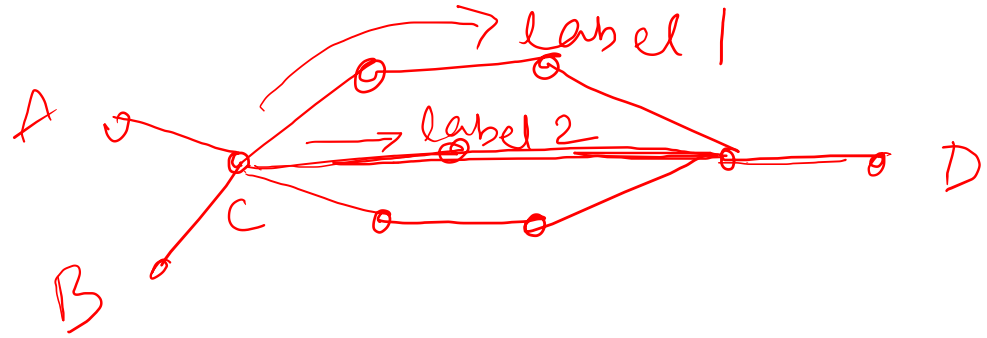


Routing protocols



- Basic idea: exchange information about IP prefixes managed by each router, construct network topology, compute shortest path
  - Notion of a link metric to help identify shortest/most optimal path
  - Route computation happens periodically, and in response to failures / network topology changes
- Link state (LS) routing protocols: each router tells the entire network about its prefixes and links, everyone knows full network topology and computes routes
- Distance vector (DV) routing protocols: each router tells its neighbors about all prefixes it knows about, routers pick which neighbor to go through for every prefix
- LS protocols have greater information to exchange but result in more accurate network topology, so faster convergence in case of failures
- Intra-domain routing protocols run between IP routers within an organization, compute routes between subnets of an AS, e.g., OSPF (Open Shortest Paths First) is LS protocol
- Inter-domain routing protocols run between border routers and compute routes across Internet, e.g., BGP (Border Gateway Protocol) is variant of DV
  - Need not be just shortest paths, but also policy considerations

Label switched routing



- Traditional routing is destination IP based shortest path routing
  - Forwarding table has shortest path for every IP prefix
  - Look up destination IP address, forward along shortest path to next IP hop
- Issues with shortest path routing
  - Some popular routes/links in a network can get very congested
  - Shortest path re-computation can take time during failures
- Alternative model employed within large ISPs and data center networks is called label switched routing, e.g., MPLS (multi protocol label switching)
  - Key idea: attach an extra label to a packet in addition to IP addresses
  - Create separate forwarding tables based on labels, can bypass shortest IP path routing within network
  - Useful for fast recovery from failures, precompute backup paths before routing converges
  - Useful for traffic engineering, can pin different flows to same destination to different network paths for load balancing traffic

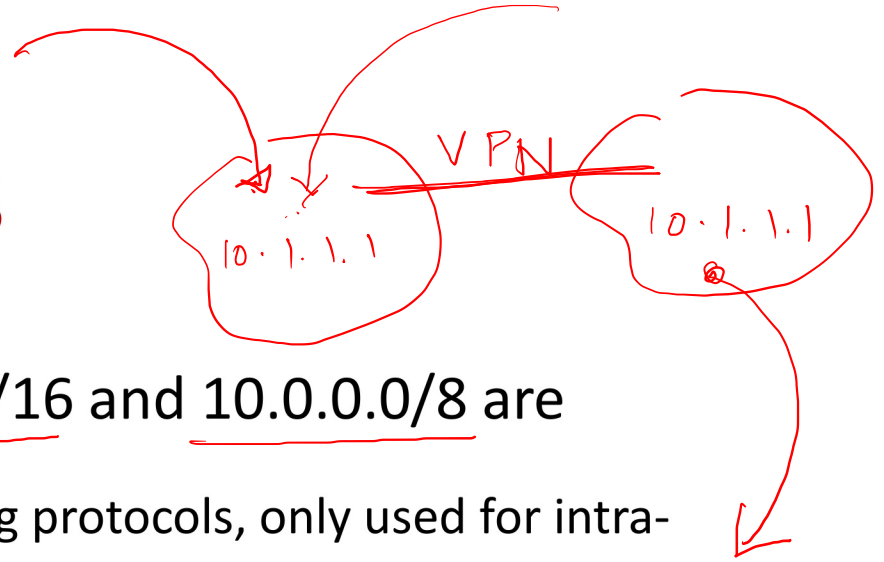
A – D label 1
B – D label 2

Software defined networking (SDN)

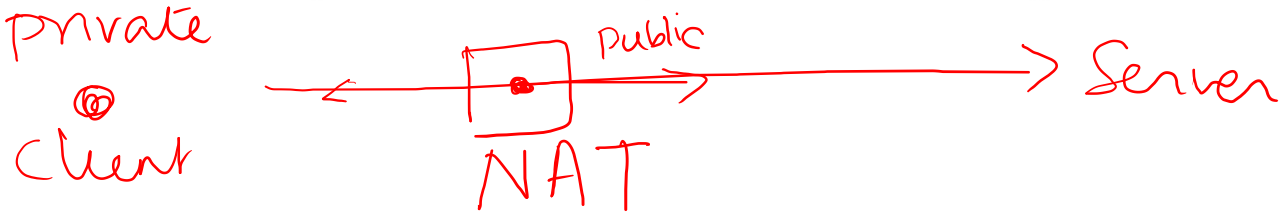
- Alternate way to design networks, useful in large datacenters or ISPs
- Key idea: separate control plane (routing) from data plane (forwarding)
  - Today, IP routers perform both routing and forwarding
  - With SDN, routing is performed in a centralized software SDN controller
  - SDN controller informs data planes nodes about how to perform forwarding using special protocols (e.g., Open Flow)

- Why SDN?
  - Can better perform routing and traffic engineering in centralized controller that has complete network visibility
  - Data plane hardware routers are simplified, complexity moved to software controller
- SDN is being used to efficiently manage large data center networks

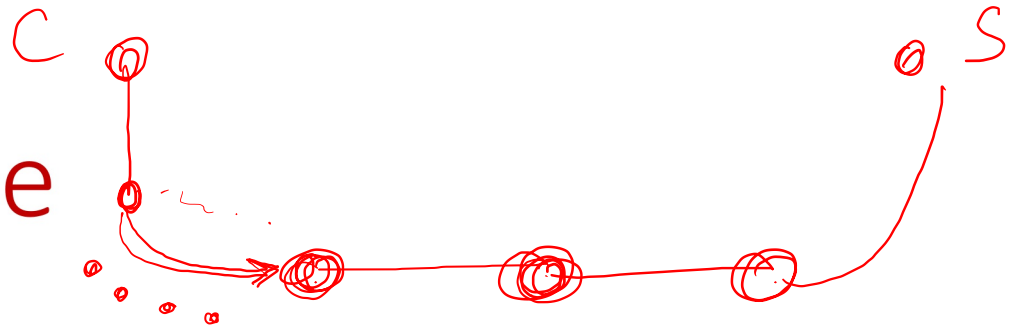
Public and private IP addresses



- Private IP addresses: two special prefixes 192.168.0.0/16 and 10.0.0.0/8 are reserved for use in organizations internally
  - These addresses are not announced via inter-domain routing protocols, only used for intra-domain routing and forwarding
  - Multiple organization can reuse same private IP addresses from private prefixes
  - Multiple islands of private IP addresses can be connected over virtual private networks (VPN), which tunnel private IP datagrams over public Internet
- Why private IP addresses?
  - IPv4 addresses are close to being exhausted, IPv6 not fully deployed
  - Isolate hosts within organization for security
- Organizations today use combination of public and private IP addresses
  - Servers that receive connections from outside clients get public IP addresses
  - Clients that contact outside servers are assigned public IP addresses temporarily via Network Address Translators (NAT), which replaces private IP with public IP when packets are leaving network and vice versa



Example: accessing a website



- A client connects to a server to access an Internet service
  - Client resolves the DNS name of server, obtains IP address
  - Client opens socket, connects it to server socket, sends message into socket
  - OS adds transport/IP/Ethernet headers, sends the packet out via the Ethernet card
  - Ethernet switches forward frame via link layer to IP router
  - Packet traverses series of IP routers till it reaches server
- What if something goes wrong? Use debugging tools
  - nslookup: resolve DNS name, check if DNS is the problem
  - ping: sends special ping packets to IP address to see if it is reachable
  - traceroute: prints out information about each IP hop along the path
  - ifconfig: information about Ethernet device
  - netstat: information about all ongoing socket connections, listen sockets, open ports

Summary

- In this lecture:
  - IP addressing, routing, forwarding
  - Link layer
  - End-to-end packet flow, debugging tools
- Learn to use various debugging tools. Use traceroute to understand the network path between your computer and any server.