Unit1_Introduction 协议栈

- **Network classification by distance**

PAN (1m-10m, e.g., room), LAN (10m-1km, e.g., building/campus), MAN (1km-100km, e.g., town/country), WAN (100km 1000km, e.g., continent), Internet (10000km, e.g., planet)

- **What is Internet?** 组成，架构，服务

Components: host/end systems + communication links + routers. **Architecture:** network of networks, loosely hierarchical, public Internet versus private intranet. **Service:** communication infrastructure, reliable/best effort data delivery.

- **Network architecture** 架构

Client/server model: client host requests, receives service from always-on server. **Peer-peer model:** minimal or no use of dedicated servers, end systems interact and run programs that perform both client and server functions.

- **Access types** 接入方式

Dial-up modem: telephony infrastructure, share physical line (surf or phone). **Digital subscriber line (DSL):** telephony infrastructure, dedicated physical line to center office. **Cable modem:** cable TV infrastructure, homes share access to router. **Ethernet:** end device + switch + router. **Wireless:** shared wireless access network, via base station/access point, e.g. Wi-Fi (802.11b/g), WiMAX (wireless interoperability for microwave access, IEEE 802.16), LTE (long term evolution)

- **Link types** 链路分类

Guided media: signals propagate in solid media, e.g., copper (铜线), fiber (光导纤维), coax (线缆). **Specific type (guided):** twisted pair (双绞线), coaxial cable (同轴电缆), fiber optic cable (光缆). **Unguided media:** signals propagate freely, e.g., radio. **Radio link types:** terrestrial microwave, LAN (e.g., Wi-Fi), wide-area (e.g., cellular), satellite.

- **Network core** 核心的架构，实现方式

Architecture: mesh of interconnected routers. **Approaches:** ①circuit switching: dedicated circuit per call. ②packet switching: data sent through net in discrete "chunks" (Reality: pure①/pure②/mixture①+②)

- **Circuit switching**

Feature: ①End-end resources (like bandwidth, switch capability) reserved for "calls" ②no sharing ③provide guaranteed service (a constant speed). **Bandwidth divide:** FDM (Frequency-Division Multiplexing), TDM (Time-Division Multiplexing)

- **Packet switching (datagram networks + virtual circuit networks)**

Feature: ①each end-end data stream divided into packets. ②each packet uses full link bandwidth ③resource contention (no admission control, congestion, store and forward). **Store and forward:** A packet (size L) transmit through a link (bandwidth R) with 2 routers in the link, need 3L/R secs (store all the packet then push out at a router). **Comments:** ①great for burst data (resource sharing, no call setup) ②excessive congestion (packet delay/loss, need protocols for reliable transfer and congestion control). **Virtual-Circuit Packet Switching (+Unit2):** ①Data is transmitted as packets. ②All packets from one packet stream are sent along a pre-established path according to VC identifier, call setup for each call before data can flow and teardown afterwards ③Packets from different virtual circuits may be interleaved ④every router on source-destination path maintains "state" for each passing connection ⑤Guarantees in-sequence delivery of packets.

- **Delay**

Nodal delay = processing + queuing + transmission (push out packet, Size/Bandwidth) + propagation **Queuing delay:** traffic intensity = (packet length * average packet arrival rate
) / (bandwidth), this value -> 1, queuing delay becomes large. **Other delay:** ①purposefully delay (determined by protocol) ②packetization delay (in Voice over-IP (VoIP) applications)

- **Why layering?**

①explicit structure allows identification, relationship of complex system's pieces. ②modularization eases maintenance, updating of system.
Lead to some problems: ①Functionality may be duplicated. ②One layer may need information present only in another layer.

- **Protocol stack**

Internet protocol stack: ③**application (message):** supporting network applications (e.g., FTP, SMTP, HTTP) ②**transport (header[port]+①->segment):** process-process data transfer (e.g., TCP, UDP) ③**network (header[ip address]+②->datagram):** routing of datagrams from source to destination (e.g., IP, routing protocols) ④**link (header[physical address]+③+tail->frame):** data transfer between neighboring network elements (e.g., PPP, Ethernet) ⑤**physical:** bits "on the wire" **OSI model:** ①... ②**presentation:** allow applications to interpret meaning of data, e.g., encryption, compression, machine-specific conventions ③**session:** synchronization, checkpointing, recovery of data exchange ④... ⑤... ⑥... ⑦...

- **About security**

Denial of service (DoS): ①an attack against any system component that attempts to force that system component to limit, or even halt, normal services. ②only from one host or network node. **Distributed denial of service (DDoS):** ①more than one attack source. ②consume the resource of target host so that normal service cannot be provided
Approach: 1(attacker) -> n (masters) -> n*m (slaves) -> 1 (target) [HW3]: Why hard to defend. **IP spoofing:** send packet with false source address

Unit2_IP Technology IPv4, IPv6

- **Key functions of network layer** 网络层功能

①forwarding: move packets from router's input to appropriate router output (IP protocol) ②routing: determine route taken by packets from source to destination (routing algorithms)

- **Virtual circuit (VC) network (+Unit1)**

Function: Provides network-layer connection service (analogous to TCP) **Different to TCP:** ①service: host-to-host according to IP address (TCP: port to port) ②no choice: network provides one or the other ③implementation: in network core and end systems. **VC:** path from source to destination + VC number + entries in forwarding tables. **VC number:** can be changed on each link [HW5: not same VC number]. **Signaling protocols:** used to setup, maintain, teardown VC in ATM

- **Datagram network**

Function: Provides network-layer connectionless service (analogous to UDP). **Feature:** ①no call setup at network layer ②no state about end-to-end connections at router ③no network-level concept of "connection" ④packets forwarded using destination host address ⑤packets between same source-destination pair may take different paths

- **Network layer protocol**

IP (Internet Protocol). **ARP** (Address Resolution Protocol): IP address -> physical address ①ARP table (IP, MAC, TTL) ②TTL: times out, delete the mapping ③broadcasts ARP query contains destination IP (MAC: FF-FF-FF-FF-FF-FF) ④destination replies MAC unicast. **RARP** (Reverse Address Resolution Protocol): physical address->IP address.

ICMP (Internet Control Message Protocol): used to communicate network-level information (error reporting e.g. unreachable; echo request/reply e.g. ping), ICMP messages carried in the data portion of IP datagrams. **IGMP** (Internet Group Management Protocol): Host uses IGMP to announce participation in multicast (more see Unit9)

- **IPv4**

Header length: 20B (32b*5) **MTU** (maximum transmission unit): largest possible link-level frame. MTU=header+data. **Fragmentation:** "reassembled" only at final destination IP header bits used to identify, order related fragments. e.g., 4000 byte datagram, MTU=1500 B ①1480B data area per fragment ②offset = 1480/8 for second fragment (0, 185, 370) ③the last fragment fragflg = 0, others =1. **Special IP:** ①all0 (startup, source) ②all0+hid (a host of this network, source) ③all1 (local/limited broadcast, destination) ④nid+all1 (directed broadcast, destination) ⑤ nid+all0 (network itself/directed broadcast, destination) ⑥127+notall0/1 (for loopback, source/destination) **Classful address schema: Advantage:** A router can keep one routing entry per network instead of per destination host. **Disadvantage:** Requiring a unique prefix for each physical network would exhaust the address space quickly as the Internet proliferates. **Solution:** unnumbered point-to-point links, proxy ARP, and subnet addressing. **Subnet:** ①device interfaces with same subnet part of IP address. ②can physically reach each other without intervening router **Classless Inter-Domain Routing (CIDR)** was invented to use address space more efficiently. Notation: a.b.c.d/x. **Longest Match:** subnet mask AND des IP -> network id, choose the longest match. **ICANN:** (Internet Corporation for Assigned Names and Numbers), only largest ISPs need to contact. **Address classification:** A: 1.0.0.0-127.0.0.0, B: 128.0.0.0-191.255.0.0, C: 192.0.0.0-223.255.255.0, D: 224.0.0.0-239.255.255.255 (for multicast), reserved: 224.0.0.0-239.255.255.255. **Private Address:** 10.0.0.0/8, 172.16.0.0/12, 192.168.0.0/16, 169.254.0.0/16.

- **NAT (Network Address Translation)**

Motivation: local network uses just one IP address as far as outside world is concerned. **Benefit:** ①simple gateway between Internet and private network ②simple security due to stateful filter implementation ③privacy and topology hiding. **Argument:** ①routers should only process up to layer 3 (but NAT provide services of transport layer) ②lead to NAT traversal problem->**Solution:** statically configure, universal plug and play (UPnP) internet gateway device (IGD) protocol and relaying (through another site)

- **IPv6**

Motivation: ①32-bit address space soon to be completely allocated (**Approach** yo slow the consumption rate: Dial-access/PPP/DHCP, Strict allocation policies, CIDR, NAT) ②helps speed processing/forwarding ③to facilitate QoS/resource allocation. **Difference:** [HW6: benefits] ①Options indicated by "Next Header" field. [tip: each header 40 Bytes, order: hop-by-hop, Routing, Fragment, Authentication, Encryption, Destination, only hop-by-hold is processed at a hop, for routing header: A send to D through B, C-> src: A, des: B, routing header: C, D]; ②Header Checksum eliminated to reduce processing time at each hop; ③Fragmentation: (a)move to extension header (b)fragmentation is end-to-end function, no fragmentation occurs in intermediate routers (c)use guaranteed minimum MTU of 1280 octets (8bits) [tip: MTU is 1280 for ipv6, 68 for ipv4, if MTU <1280, use link-specific fragmentation at end device] or perform **Path MTU Discovery** [tip: send a specific packet, ICMP "packet too big" message would occur if packet is too big] to identify the minimum MTU along the path to the destination. [**Problem:** routes cannot be changed as easily as those in IPv4 because a change in a route can also change the path MTU] ④TTL->Hop Limit ⑤**Protocol**->Next Header; ⑥**Service Type**->Traffic Class ⑦Addresses increased 32 bits-> 128 bits ⑧Flow Label added, identify datagrams in same "flow" (e.g., two applications that need to send video can establish a flow on which QoS is guaranteed). [HW: Private addresses]. **Transition(过渡):** IPv6 deployments will occur piecewise from the edge. **Co-Existence(共存) Techniques:** ①dual-stack techniques: to allow IPv4 and IPv6 to co-exist in the same devices and networks ②tunneling techniques: IPv6 carried as payload in IPv4 datagram among IPv4 routers ③translation techniques: to allow IPv6-only devices to communicate with IPv4-only devices

Unit3_Network Switching

- **Switching**

Switch type: ①memory ②bus ③cross bus.. **Output port queueing:** buffering when arrival rate via switch exceeds output line speed. **Input Port Queuing:** fabric slower than input ports combined [HW8: queue]

- **Devices [HW2: comparison]**

Hubs: no buffer, no filtering, no redirection, no CSMA/CD **Switch:** store forward, no collisions, full duplex, network is restricted to a spanning tree in order to prevent the cycling of broadcast storm, maintain switch tables, implement filtering, learning algorithms **Bridge:** only has one incoming and one outgoing port, perform in software (switch hardware) **Router:** provide firewall protection and allow the network to be built with a rich topology, maintain routing tables, implement routing algorithms.

- **[HW11: IP over ATM/IP over SDH/IP over WDM]**
- **[HW9&10: MPLS]**

Unit4_Transport Layer TCP, UDP, GBN, SR

- **Transport-layer protocols**

TCP (transmission control protocol): Header length: 20B, reliable, in-order delivery, congestion control, flow control, connection-oriented, integrity checking. **UDP (user datagram protocol):** ①unreliable, unordered delivery, error checking. ②Header length: 8B ③used for streaming multimedia apps (loss tolerant, rate sensitive) ④DNS, SNMP

TCP & UDP Segment structure

- **Pipelining Protocols [HW16: Compare GBN/SR/TCP]**

GBN (go-back-N/sliding window protocol): receiver only send cumulative ACKs, drop unexpected packets; sender sets timer for oldest unACKed packet, and will retransmit a series packets if a former packet is lost. **SR (selective repeat):** receiver buffers and ACK each packets, sender sets timer for each unACKed packet, only retransmit packets which are in error. (Requirement: window size <= half of seq # size, if not, can't distinguish new packet and retransmission)

- **TCP [HW16: Compare GBN/SR/TCP]**

Seq & ACK (Telnet): initial number (given or random) for client and server (seq1, seq2). ACK for expect segment seq (rcvseq+rcvdata size). [e.g., client (seq=42, ACK=79, data=C), server (seq=79, ACK=43, data=C), client (seq=43, ACK=80), two C for echo back). **Fast retransmit:** if sender receives 3 same ACKs, resend segment before timer expires. **Three way handshake:** ①client host sends TCP SYN segment to server (SYN=1, seq=c_isn) ②server host receives SYN, replies with SYNACK segment (SYN=1, seq=s_isn, ACK=c_isn+1) ③client receives SYNACK, replies with ACK segment (SYN=0, seq=c_isn+1, ACK = s_isn+1) **Close connection:** ①client send FIN ②server ACK ③server send FIN ④client ACK. **Flow control:** receiver send its rcv window size in the segment back to sender.

- **Congestion control [HW17: Compare flow/congestion]**

Approach: ①End-end: end-system observed loss, delay; approach taken by TCP ②Network-assisted: routers provide feedback to end systems (e.g. IBM SNA, DECbit, TCP/IP ECN and ATM)

- **TCP congestion control**

Feature: AIMD (additive increase, multiplicative decrease): increase cwnd by 1 MSS every RTT until loss detected (CA mode), cut cwnd in half after loss [tip: sending rate=cwnd/RTT]. **TCP congestion control:** ①When cwnd is below Threshold, sender in **slow-start** (double cwnd every RTT) phase, window grows exponentially. ②When cwnd is above Threshold, sender is in CA (congestion-avoidance) (cwnd++ every RTT) phase, window grows linearly. ③When a triple duplicate ACK occurs, Threshold set to cwnd/2 and cwnd set to Threshold. ④When timeout occurs, Threshold set to cwnd/2 and cwnd set to 1 MSS. **Tail-drop policy** cause **global synchronization [HW20], reason:** under a tail-drop policy, the router will discard one segment from N connections rather than N segments from one connection, the simultaneous loss causes all N instances of TCP to enter slow-start at the same time and throughput decreases suddenly, after the network recovers, throughput will suddenly increase a lot. **Solution->RED (Random early detection):** instead of waiting until the queue overflows, a router slowly and randomly drops datagrams as congestion increases. **Throughput** = (W/RTT)*(1+1/2)/2=0.75 W/RTT (W is window size when loss occurs) **Fairness:** for idealized conditions (same MSS and RTT), TCP is fair. In practice, those sessions with a smaller RTT are able to grab the available bandwidth at that link more quickly as the link becomes free. Moreover, consider UDP and parallel TCP connections. **Explicit feedback mechanisms:** selective acknowledgement (SACK), explicit congestion notification (ECN)

Unit5_MultimediaQoS 区分服务模型

- **Multimedia applications (delay sensitive, loss tolerant)**

①**stored streaming:** e.g., YouTube, media stored at source, transmitted to client, client playout begins before all data has arrived. Constraint: in time for playout ②**live streaming:** e.g., IPTV, can't fast forward ③**real time interactive:** e.g. IP telephony, video conference, distributed interactive worlds. **Evolve to better support multimedia Approach:** ①Integrated services philosophy: fundamental changes in Internet ②Differentiated services philosophy, fewer changes to Internet infrastructure ③Laissez-faire: no major changes, provide more bandwidth when needed (e.g. CDN [HW23], application-layer multicast [tip: **Hard guarantee:** receive QoS with certainty. **Soft guarantee:** with high probability]

- **Supporting Multimedia applications**

(Approach, Unit of allocation, Guarantee, Mechanisms) ①**making the best of best-effort service,** none, none or soft, application layer support, CDN, over-provisioning ②**Differential QoS,** classes of flows, none or soft, policing, scheduling ③**guaranteed QoS,** individual flows, soft or hard, once a flow is admitted, policing, scheduling, call admission and signaling.

- **Streaming Multimedia: UDP or TCP?**

UDP: sends at rate appropriate, send rate=encoding rate=constant rate, fill rate=constant rate-packet loss. **TCP:** send at maximum possible rate, fill rate fluctuates due to TCP congestion control, HTTP/TCP passes more easily through firewalls. [Tip: **Handle different client receive rate capabilities:** server stores, transmits multiple copies of video, encoded at different rates]

- **Principles for QoS Guarantees**

①packet classification ②isolation: scheduling and policing ③high resource utilization ④call admission

- **Scheduling Mechanisms**

①FIFO ②Priority ③Round robin scheduling / fair queuing: cyclically scan class queues ④Weighted Fair Queuing (WFQ): each class gets weighted amount of service in each cycle

- **Policing Mechanisms [HW25]**

Token Bucket: bucket can hold b tokens, tokens generated at rate r token/sec unless bucket full, the max number of packets be sent in a given time T is (r*T+b), if scheduling is WFQ, the max delay is (b/R*w0/sum(wi)).

- **Differentiated services [HW21]**

Edge router: per-flow traffic management, packet classification and traffic conditioning, marks packets as in-profile and out-profile. Core router: per class traffic management, buffering and scheduling based on marking at edge, forwarding, preference given to in-profile packets. [tip1: packet is marked in the 8-bit (6 bits used for Differentiated Service Code Point (DSCP) determine Per-Hop Behavior (PHB), 2 bits not used) Type of Service (**TOS**) in IPv4, and 8-bit Traffic Class in IPv6] [PHB result in a different observable (measurable) forwarding performance behavior, PHB does not specify what mechanisms to use to ensure required PHB performance behavior. Two type: **Expedited Forwarding (Premium):** pkt departure rate of a class equals or exceeds specified rate, **Assured Forwarding:** define 4 classes of traffic]

- **Integrated Services [HW21]**
- **RSVP [HW24]**

Unit6_Internal Routing 路由算法 LS,DV,RIP,OSPF

- **Routing algorithms classification**

①**Global:** all routers have complete topology, link cost information, e.g., LS. **Decentralized:** router knows physically-connected neighbors, link costs to neighbors, iterative process of computation, exchange of information with neighbors, e.g. DV ②**Static:** change slowly, only for simplest cases. **Dynamic:** periodic update in response to topology or link cost changes, necessary in large internets. ③**Load-sensitive** or **load-insensitive** (today's algorithms)

- **LS (Link state/Dijkstra) [HW26]**

Complexity: with n nodes, E links, O(nE) msgs sent, n(n+1)/2 comparisons: O(n^2), possible: O(nlogn). **Oscillations possible [HW28] .Robustness(鲁棒性):** node can advertise incorrect link cost, each node computes only its own table, somewhat separated route calculations providing a degree of robustness

- **DV (Distance vector) [HW26]**

May be routing loops, **"count to infinity" problem [HW29], Solution: Poisoned reverse: [HW30]** to avoid routing loops, when a router find a subnet is not alive, it will set the cost infinite (e.g. 16) when broadcast to other routers instead of deleting it immediately.

- **RIP (Routing Information Protocol) [HW27: Compare]**

Basic idea: ①distance vectors: exchanged among neighbors every 30 sec via Response Message ②each advertisement: list of up to 25 destination subnets within AS. Timer: 30s for routing-update, 180s for time out, 120s for garbage collection (delete route)
Disadvantages: [HW31]

- **OSPF (Open Shortest Path First) [HW27: Compare]**

Scale: 150~500 routers/area. **Basic idea:** ①Distributed, replicated database model (Each router builds a topology database describes complete routing topology) ②Link state database (identical for all the routers) ③LSA (Information about adjacencies sent to all routers) ④A "shortest path" algorithm is used to find best route (dijsktra) (Converge as quickly as databases can be updated, every router calculate itself routing table independently) **Two-level hierarchy:** local area, backbone. **Link state advertisement (LSA)** is bounded by area. **Advantage:** security, load balancing, type of service (TOS) routing, integrated unicast and multicast support, hierarchical.

Unit7_External Routing 区域间选路 BGP，热土豆，冷土豆

IGP (interior gateway protocol): OSPF, IS-IS, RIP, EIGRP (cisco). **EGP (exterior gateway protocol):** BGP. **IXP (internet exchange point):** a physical infrastructure through which ISPs and CDNs exchange Internet traffic between their networks. **ASN (autonomous system number):** 自治系统号

- **Why do we need EGP? [HW33: IntraAS/InterAS routing]**

①Scalability (hierarchy, limit scope of failure) ②Flexibility in choosing routes ③Define administrative boundary ④Policy (control reachability to prefixes)

- **Interconnections type**

①**Transit** ②**Peering** If A peer with B, B peer with C, A's customer could not send data to C directly. [HW34]

- **BGP (Border Gateway Protocol)**

Layer: use reliable transport i.e., TCP **Function:** ①Obtain subnet reachability information from neighboring ASs. ②Propagate reachability information to all AS-internal routers. ③Determine "good" routes to subnets based on reachability information and policy. **Neighbor relationships:** ①**eBGP** session spans two ASs, to share connectivity information across AS Network Layer Reachability Information (NLRI). ②**iBGP** session between routers in the same AS, carrying information within an AS. **Handle prefix:** because of longest match principle, if AS has 3 subnets, 138.16.64/24, 138.16.65/24 and 138.16.66/24, it will aggregate prefixes to 138.16.64/22 in its advertisement, because another AS will advertise 138.16.67/24. **AS-PATH** contains ASs through which prefix advertisement has passed, e.g., AS2 receive prefix from AS1, when AS2 advertise to AS3, AS-PATH=AS2 AS1. To **prevent loop**, AS will never accept a route containing AS itself. **NEXT-HOP** is the router interface that begins the AS-PATH and indicates specific internal-AS router to next-hop AS. Every time a route announcement crosses an AS boundary, the Next Hop attribute is changed to the IP address of the border router (when the announcement is in an AS, the Next-HOP not change). **Route selection:** ①local preference value attribute: policy decision up to AS's network administrator (Highest values are selected) ②shortest AS-PATH (Distance vector algorithm; Distance metric: # AS hops, NOT # router hops) ③closest NEXT-HOP router/hot potato routing (Least-cost path determined by intra-AS algorithm) ④additional criteria (can be more complicated)
Hot potato routing & Cold potato routing [HW35&36]

Unit8_Virtual Private Networks (VPN)

Goal: to keep internal datagrams private while still allowing external communication **Main benefit:** reduce cost. **Other benefit: Scalability NAS (Network Access Servers):** a device that interfaces between an access network and a packet-switched network, serve as a tunnel endpoint in a remote access VPN.

- **Types**

①**Site-to-site:** allow connectivity between an organization's (or organizations') geographically dispersed sites (such as a head office and branch offices). Two types of site-to-site: **Intranet:** Allow connectivity between sites of a single organization. **Extranet:** Allow connectivity between organizations such as business partners or a business and its customers. ②**Remote access:** allow mobile or home-based users to access an organization's resources remotely.

- **Protocols for site-to-site [HW37-41]**

①IPsec (IP security) ②GRE (Generic Routing Encapsulation) ③L2TPv3 (Layer Two Tunneling Protocol version 3) ④Q-in-Q (IEEE 802.1Q tunneling) ⑤MPLS (multiprotocol label switching) LSPs

- **Protocols for remote access [HW37-41]**

①L2F (The Layer Two Forwarding) Protocol ②PPTP (The Point-to-Point Tunneling Protocol) ③L2TPv2/L2TPv3 ④IPsec ⑤SSL (Secure Sockets Layer)

- **Most popular:** PPTP, L2TP and IPsec
- **Protocol Classified by layer**

①Application to Application: SSL ②End to End: IPSec Transport Mode ③Gateway to Gateway: PPTP, L2TP/IPSec, IPSec Tunnel Mode ④Client to Gateway: L2TP/IPSec

- **VPN Critical Functions**

①Authentication: validates that the data was sent from the sender. ②Access control: limiting unauthorized users from accessing the network. ③Confidentiality: preventing the data to be read or copied as the data is being transported. ④Data Integrity: ensuring that the data has not been altered **SA:** Before sending IPsec datagrams from source entity to destination entity, the source and destination entities create a network-layer logical connection – called a **security association**

Unit9_Multicasting 多播，生成树

- **In-network duplication**

①**Uncontrolled flooding:** when node receives broadcast packet, sends copy to all neighbors (except the source neighbor) Problem: cycles, broadcast storms. ②**Controlled flooding:** node only broadcast packet if it hasn't broadcast same packet before. Approach: sequence-number-controlled / reverse path forwarding (RPF) / reverse path broadcast (RPB) ③**Spanning tree:** no redundant packets received by any node [tip: a node need not be aware of the entire tree, simply needs to know its spanning-tree neighbors.]

- **Multicast:** one sender to many receivers

Control scope: ①IP's TTL (Time-To-Live) field ②Administrative scoping. **Local->IGMP:** ①When it joins a group, host sends message (REPORT) declaring membership. ②Multicast router periodically polls (QUERY) a host to determine if any host on the network is still a member of a group **Wide area (among routers)->multicast trees:** local router interacts with other routers to receive multicast datagram flow

- **Approaches for building multicast trees**

①**Source-based tree:** one tree per source, e.g., shortest path trees [MOSPF (Multicast extensions to OSPF)], reverse path forwarding [DVMRP (Distance-Vector Multicast Routing Protocol), PIM-DM (Protocol Independent Multicast-Dense Mode)] ②**Group-shared tree:** group uses one tree, e.g., minimal spanning [Steiner], center-based trees [CBT (Core-Based Trees), PIM-SM (Protocol Independent Multicast-Sparse Mode)]

Homework

1. 35users, active 10% of time, probability > 10 active at same time
P = 1-[sum n from 0 to 10 (C(n,35) * 0.9^(35-n) * 0.1^n)]

The benefits of IPv6 when compared with IPv4: Larger Addresses +Extended Address Hierarchy + Flexible Header Format + Improved Options + Support for Autoconfiguration and Renumbering + Support for Resource Allocation

Compare and contrast the IPv4 and the IPv6 header fields

Streamlined: fragmentation fields moved out of base header + IP options moved out of base header, indicated by "Next Header" field + Header Checksum eliminated to reduce processing time at each hop + Header Length field eliminated + Length field excludes IPv6 base header / Revised: Time to Live -> Hop Limit + Protocol -> Next Header + Service Type -> Traffic Class + Addresses increased 32 bits -> 128 bits / Extended: Flow Label field added, identify datagrams in same "flow."

Give more detailed information about the plane of network layer:

Forwarding(in Data plane):
move packets from router's input to appropriate router output
- IP protocol
- involves a single router
- forwarding takes place typically in a few nanoseconds, and thus is typically implemented in hardware.

forwarding: process of getting through single interchange

Routing(in Control plane):
determine route taken by packets from source to dest. (network-wide process)
- routing algorithms
- involves all of a network's routers
- routing takes place on much longer timescale (typically seconds) and often is implemented in software.

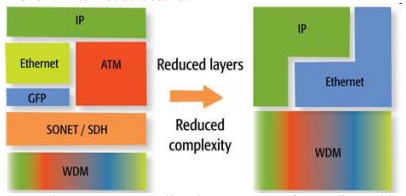
routing: process of planning trip from source to dest.

Queue: Output port: buffering when arrival rate via switch exceeds output line speed + queuing (delay) and loss due to output port buffer overflow! + Consequence: a packet scheduler at the output port must choose one packet, e.g., selection can be based on first-come-first-served (FCFS) scheduling / Input port: Fabric slower than input ports combined -> queuing may occur at input queues + Head-of-the-Line (HOL) blocking: queued datagram at front of queue prevents others in queue from moving forward + queuing delay and loss due to input buffer overflow!

Switching Interconnection: Overcome bus bandwidth limitations + A crossbar switch is an interconnection network consisting of 2n buses that connect n input ports to n output ports. + Advanced design: fragmenting datagram into fixed length cells, switch cells through the fabric. + Cisco 12000: switches 60 Gbps through the interconnection network + The fastest switching type

Present router: run routing algorithms/protocol (RIP, OSPF, BGP) + forwarding datagrams from incoming to outgoing link

IP-over / Internet structure:



SDN: The remote controller that computes forwarding tables and interacts with routers is implemented in software + providing many of these network-layer functions and certain link-layer functions as well, in a modern, elegant, and grated manner

Load balancer: Inside the data center, the external requests are first directed to a load balancer whose job is distributing requests to the hosts, balancing the load across the hosts as a function of their current load + a large data center will often have several load balancers, each one devoted to a set of specific cloud applications + Such a load balancer is sometimes referred to as a "layer-4 switch" + since it makes decisions based on the destination port number (layer 4) as well as destination IP address in the packet. + the load balancer not only balances the workload across hosts, but also provides a NAT-like function.

limited host-to-host capacity: Problem: suppose each host connects to its TOR switch with 1Gbps link rate, but the links between switches are 10 Gbps rate. If there are many flows between different racks at the same time, the max rate for two hosts may be less than 1Gbps because sharing 10 Gbps link. + Solution: deploy higher-rate switches and routers and deploy new interconnection architectures and network protocols, e.g., a fully connected topology.

Trends in data center: deploy new interconnection architectures and network protocols + employ shipping container-base modular data centers (MDCs)

MPLS applications: Traffic Engineering (use QoS to control the rate of traffic on each path) + Class of Service (Support for differentiated services), VPN (MPLS allows ISPs to offer VPN services by providing a simple, flexible, and powerful tunneling mechanism)

Compare flow control and congestion control: Flow control: end-to-end + control and handle the traffic between a sender and a receiver + slide window congestion control / Congestion control: entire network + control congestion in the network and prevent loss of packets and delay + make sure the entire work can handle the traffic that is coming to the network + slow start + congestion avoidance + fast retransmit + fast recovery

TCP = GBN / SR: hybrid of GBN and SR protocols / GBN-like: TCP sender need only maintain the smallest sequence number of a transmitted but unacknowledged byte (SendBase) and the sequence number of the new byte to be sent (NextSeqNum) / SR-like: Many TCP implementations will buffer correctly receive but out-of-order segments. TCP, on the other hand, would retransmit at most one segment, namely, segment n. Moreover, TCP would not even retransmit segment n if the acknowledge for segment n+1 arrived before the timeout for segment n. A proposed modification to TCP, the so-called selective acknowledgement, allows TCP receivers the acknowledge out-of-order segments selectively rather than just cumulatively acknowledge the last correctly received, in order segment.

Compare UDP and TCP: point-to-point + reliable, in-order byte stream + pipelined + send & receive buffers + full duplex data + connection-oriented + flow controlled / UDP: no frills + best effort service + connectionless + no congestion control + loss tolerant + rate sensitive

Compare three service models: Best effort service model: single model is a single service model try its best to send message but does not provide guarantee for the performance like time delay, reliability + no QoS, simple, all packets are equal at the router, no special treatment for any delay-sensitive multimedia applications / Intserv (Integrated Services): Reserved Resources + Call Setup / call admission + architecture for providing QoS guarantees in IP networks for individual application sessions / Diffserv (Differentiated Services): aims to handle different "classes" of traffic in a scalable and flexible manner + scalability: simple functions in network core, relatively complex functions at edge routers (or hosts) + flexibility: don't define specific service classes, but provide functional components to build service classes

VoIP: a methodology and group of technologies for the delivery of voice communications and multimedia sessions over IP networks: Receiver -> analogy-digital converter -> compression encoder -> IP encapsulation -> digital-analog converter ->player

CDN: replicate stored content and put the replicated content at the edges of the Internet. CDNs provide a differentiated service to content providers + replicate content at hundreds of servers throughout Internet + placing content "close" to user + CDN server typically in edge/access network + Servers nearest to the website visitor respond to the request. The content delivery network copies the pages of a website to a network of servers that are dispersed at geographically different locations, caching the contents of the page. When a user requests a webpage that is part of a content delivery network, the CDN will redirect the request from the originating site's server to a server in the CDN that is closest to the user and deliver the cached content. CDNs will also communicate with the originating server to deliver any content that has not been previously cached

Token bucket / leaky bucket: Token Bucket: limit input to specified Burst Size and Average Rate and can allow a certain degree of burst transmission/ Leaky bucket: limit the transmission rate

Peering and transit: Peering: peering is a business relationship whereby two companies interconnect directly without charging, RECIPROCALLY exchange access to each other's customers + Peering is open only to traffic coming from a peer's end-users or from networks that have bought transit. / Transit Provider sells metered access to the Global Internet + A transit provider will not announce its peering and transit

Compare LS and DV: LS: Dijkstra's algorithm+ all routers have complete topology, link cost info + net topology, link costs known to all nodes + computes least cost paths from one node ("source") to all other nodes / DV: Bellman-Ford Equation + router knows physically-connected neighbors, link costs to neighbors + iterative process of computation, exchange of info with neighbors + a node gradually calculates the least-cost path to a destination or set of destinations + From time-to-time, each node sends its own distance vector estimate to neighbors + Asynchronous

Compare RIP, OSPF and IS-IS: Routing Information Protocol (RIP): Routing Information Protocol (RIP): distance vector algorithm + distance metric: # of hops (max = 15 hops) + distance vectors: exchanged among neighbors every 30 sec via Response Message (also called advertisement) + UDP + slow response to change + no security + no hierarchy / Open Shortest Path First (OSPF): link state algorithm, criterion of routing: bandwidth & delay, advertisements disseminated to entire AS (via flooding), use flooding to send msgs, use IP (not UDP) directly to transmit IP datagrams + security + hierarchy / Intermediate System to Intermediate System is a routing protocol designed to move information efficiently within a computer network, a group of physically connected computers or similar devices. It accomplishes this by determining the best route for data through a Packet switching network.

Link-state algorithm (oscillations): Oscillations possible (with congestion-sensitive routing): link A and B is empty first, sender choose link A send first, then link A will become busy and if sender would send data immediately, it would choose link B + e.g., link cost = amount of carried traffic + => link cost is asymmetric + i.e., c(u,v) = c(v,u) only if the load carried on both directions on the link (u,v) is the same

Count-to-infinity: Slow convergence problem + good news travels fast + bad news travels slowly: It occurs when one router feeds another old information, which continues to propagate through the network toward infinity. This occurs when a link is removed.

Poisoned reverse: If Z routes through Y to get to X: Z tells Y its (Z's) distance to X is infinite (so Y won't route to X via Z) + This ensures all routers in the domain receive the poisoned router update. + used to prevent ping-pong loops (infinite distance = 16 hops)

BGP incidents: illegitimate takeover of groups of IP addresses by corrupting Internet routing tables maintained using the Border Gateway Protocol (BGP) + Like the TCP reset attack, session hijacking involves intrusion into an ongoing BGP session, i.e., the attacker successfully masquerades as one of the peers in a BGP session, and requires the same information needed to accomplish the reset attack

Compare Intra- and Inter-AS routing: Inter-AS: Policy based + The Routing Domain of BGP is the entire Internet Used to convey routing information between ASes / Intra-AS: no policy decisions needed focus on performance + Metric based + Automatic discovery + Generally trust your IGP routers + Routes go to all IGP routers

Hot/cold potato routing: With hot potato routing, you want the traffic going to a particular destination off your network ASAP + the practice of handing over traffic at the earliest convenience + Go for the Closest Egress Point + minimize the amount of works thus resulting in lower QoS / Cold potato routing is the opposite. You want to keep that traffic on your network as long as possible transport it to a point as close as possible to the final destination before handing it over to a peer/provider (if necessary) + where you hold onto traffic as long as you can before handing it over to another network + more expensive to do and requires a level of trust between two networks that either side will not attempt to "cheat" the other

SSL / TLS: Secure Sockets Layer (SSL) — a security protocol that was originally developed by Netscape Communications, later developed into Transport Layer Security (TLS), an IETF standard, is similar to SSLv3. + usually implemented on top of any of the Transport Layer protocols + sometimes referred to as web VPNs or clientless VPNs because no special client software is required other than a web browser + More functionality can be added by installing SSL VPN client software on remote access client devices

VPN disadvantages: VPNs require an in-depth understanding of public network security issues and proper deployment of precautions + Availability and performance depends on factors largely outside of their control + Immature standards + VPNs need to accommodate protocols other than IP and existing internal network technology

Symmetric encryption: Alice's , Bob's keys are identical and are secret

Public key: A pair of keys is used + One of the keys is known to both Bob and Alice (indeed, it is known to the whole world) + The other key is known only by either Bob or Alice (but not both).

RSVP: Resource Reservation Protocol + known as a soft-state protocol, i.e., can expire + used to install state (bandwidth reservations) in routers + To implement RSVP, RSVP software must be present in the receivers, senders, and routers along the end-end path

MPLS (Operation): works by prefixing packets with MPLS header, containing one or more labels, called label stack. MPLS-labeled packets are switched after label lookup (switch instead of lookup into IP table + LSR receives a packet, it uses label included as an index to determine the next hop on the LSP and a corresponding label for packet from lookup table. The old label is removed from the header and replaced with new label before the packet routed + When forwarding IP datagram into MPLS domain, an LER uses routing information to determine label to be affixed and forwards the labelled packet into the MPLS domain. Upon receiving a labelled packet destined to exit MPLS domain, LER strips off the label and forwards the IP packet using IP forwarding rules / MPLS allows most packets to be forwarded at Layer 2 (the switching level) rather than having to be passed up 17. to Layer 3 (the routing level). Each packet gets labeled on entry into the service provider's network by the ingress router. All the subsequent routing switches perform packet forwarding based only on those labels—they never look as far as the IP header. Finally, the egress router removes the label(s) and forwards the original IP packet toward its final destination. The label determines which pre-determined path the packet will follow. The paths, which are called label-switched paths (LSPs), allow service providers to decide ahead of time what will be the best way for certain types of traffic to flow within a private or public network.

Ipsec: Symmetric Key Encryption + for securing the network-layer transport + designed to protect IP traffic between security gateways or hosts as it transits an intervening network + As well as enabling site-to-site VPNs, IPsec can also be used to securely tunnel data traffic between remote access or mobile users and a VPN gateway/concentrator

L2TPv3: Layer Two Tunneling Protocol version 3 (L2TPv3) — allows the point-to-point transport of protocols over an IP or other backbone + L2TP has limited intrinsic security, and so L2TP tunnels are often protected using IPsec + allows the point-to-point transport of protocols over an IP or other backbone

GRE: construct tunnels and transport multiprotocol traffic between CE devices in a VPN. GRE has little or no inherent security, but GRE tunnels can be protected using Ipsec + tunning protocol that encapsulates network layer inside virtual p2p links + support multiprotocol and multicast + support multipoint tunnel + provide QoS

PPTP: Point-to-Point Tunneling Protocol + layer 2 protocol hat encapsulates PPP frames in IP datagrams for transmission over an IP internetwork + lower transmission cost + lower hardware cost + lower administrative overhead + security

AIMD: The approach taken is to increase the transmission rate (window size), probing for usable bandwidth, until loss occurs. The policy of additive increase may, for instance, increase the congestion window by a fixed amount every round trip time. When congestion is detected, the transmitter decreases the transmission rate by a multiplicative factor; for example, cut the congestion window in half after loss. The result is a saw-tooth behavior that represents the probe for bandwidth

Delay Jitter: In computer networking, packet delay variation (PDV) is the difference in end-to-end one-way delay between selected packets in a flow with any lost packets being ignored. The effect is sometimes referred to as jitter

UDP socket: identified by two-tuple + (dressIP address, dest port number)

CSMA/CD: Carrier Sense Multiple Access / Collision Detection

IGMP : (Internet Group Management Protocol) announce participation in multicast. 2. Phases: When it joins a group, host sends message declaring membership; Multicast router periodically polls a host to determine if any host on the network is still a member of a group (no explicit when leaving) + R joins to group 224.2.0.1--R sends IGMP Membership-report to 224.2.0.1;DR receives it. DR will start forwarding packets for224.2.0.1 to network A; DR periodically sends IGMP Membership-Query to 224.0.0.1(all systems mcast); R answers IGMP Membership-report to 224.2.0.1

Advantages and Disadvantages of the Original Classful IP Addressing scheme: Advantage: a router can keep one routing entry per network instead of per destination host + Use classful addressing to determine the boundary between prefix and suffix, e.g., Class A partitioned an address into 8-bit network portion and a 24-bit host portion / Weakness: requiring a unique prefix for each physical network would exhaust the address space quickly as the Internet proliferates.

Mobility comparison between GSM and Mobile IP: both have high mobility and mobile user can maintain connections through multiple access point, both use indirect routing to communicate with users; Diff: mobile IP prefers user who move infrequently and can stay for a relatively long period of time because of the considerable overhead during the transmission of data; In GSM, Mobile Switching Center work instead of routers in IP network. HLR(Home Location Register), VLR(Visitor Location Register)are used to store phone num, like the IP address in IP network.

The same VC number: Replacing the number from link to link to reduce the length of the VC field in header + Permitting a different VC number for each link along the path of the VC to simplified a network management function and VC setup because each link chooses VC num independently and common VC num costs a lot

DDOS: hard to monitored or tracked, attackers hide themselves well + based on legal packets So firewall spends high-intensive check to prevent + systems optimization and increasing bandwidth cost lot, but escalation of DDoS attack costs les

Reliable data transfer (rdt): ack + retransmission + timeout + sliding window + stop-and-wait + pipelined protocols(buffering at sender and/or receiver; error recovery protocols like go-back-n, selective repeat)s

Most common VPN protocols: PPTP (point to point tunneling protocol) + L2TP + IPSES

The disadvantages of RIP: Increased network traffic: RIP checks with its neighboring routers every 30 seconds, which increases network traffic. Maximum hop count: RIP has a maximum hop count of 15, which means that on large networks, other remote routers may not be able to be reached. Closest may not be shortest: Choosing the closest path by hop count does not necessarily mean that the fastest route was selected. RIP does not consider other factors when calculating best path. RIP only updates neighbors so the updates for non-neighboring routers are not first-hand information

Overlay VPNs a VC or tunnel connects CE devices, no routing information is exchanged with the service provider (PE devices) Examples: those built using Frame Relay or ATM virtual circuits, as well as those built using GRE or IPsec tunnels

Peer VPNs PE devices are aware of customer network addressing and route customer data traffic according to customer network addressing Example: BGP/MPLS (RFC4364/2547bis) VPNs

Three basic IPv6 address types: unicast (destination address specifies a single computer, delivery to single), anycast, multicast (destination is a set of computers, possibly at multiple locations. Delivery to each member in the set using hardware multicast or broadcast if viable)

Three types of switching fabrics: twisted pair, memory, bus, crossbar

RFC: request for comments + never change once published + not all RFCs are standards

SLA: Service Level Agreement; An SLA is a formal negotiated agreement

Switch: link-layer device: smarter than hubs, take active role; transparent: hosts are unaware of presence of switches; self-learning: switches do not need to be configured

Compare switch and bridge: number of network segments, bridge 1-1 + switch 1-N + switches perform in hardware, bridges perform in software + Both store-and-forward devices; Both maintain tables topology: switch--> a spanning tree, routers--> a rich topology

MTU: maximum transmission unit + a link's maximum transmission unit transmitted over the link

Protocols: protocols define format, order of msgs sent and received among network entities, and actions taken on msg transmission & receipt

Compare Transport and network layer network layer: logical communication between hosts; transport: between processes

VC implementation: VC consists of path, VC numbers, entries in forwarding table + Replacing the number from link to link reduces the length of the VC field in the packet header. 2. VC setup is considerably simplified by permitting a different number at each link along the path of the VC. Each link can choose a VC number independently and common VC number costs a lot

Core-based trees (CBT): better for sparse network; Protocol Independent Multicast (PIM), no dependent on any specific underlying unicast routing alg. PIM-SM (like CBT), PIM-DM (use flooding to forward data)

Joining a mcast group: two-step process: local: host informs local mcast router of desire to join group (IGMP); wide area: local router interacts with other routers to receive mcast datagram flow.

QoS for networked applications: packet classification + isolation scheduling and Policing + high resource utilization + call admission + queuing delay + loss due to input buffer overflow