Evaluation Method

- Interim and Final Report

- Attendance is not Checked, but, ...

- Questions or Comments are Mandated

  – In the quarter, questions or comments with technical content must be made at least twice during lecture (may be in Japanese)

  – Good questions and comments will be awarded with points

  – Declare your name and student ID, if you make questions or comments

Evaluation with Zoom

- questions/comments should be asked/made by oral interruption (not by chat)
  - raising hand by zoom is hard to be noticed unless dedicated chair is assigned
  - don't hesitate to interrupt my talk
    - questions/comments over chat is too easy
- name/ID and points are declared and given through chat
  - use private chat, if you don't want your ID publicly viewed

For Better Verbal Communication with Zoom

- echo cancellation of zoom is, seemingly, not very good
- it is strongly recommended to turn off speakers and use head/ear phones (should be available at 100-yen shops)

Remaining Topics and Rescheduling

- the following topic will be omitted
  - 9. Routing: Traffic Engineering, ROLC, MPLS
- course survey is planned on 7/30
  - URL for the survey will be announced by chat during the lecture

Advanced Lecture on Internet Infrastructure
10. Routing: Multicast, Aggregation of Routing Table

Masataka Ohta

mohta@necom830.hpcl.titech.ac.jp

ftp://chacha.hpcl.titech.ac.jp/infra10e.ppt

What is Multicast?

- one to many, many to many communication by copying data in network
  - "broadcast" by network

- necessary in network not possible by end
  - copying end is called reflector or, IMHO improperly, application layer multicast

- intimately related to resource reservation
  - cannot adjust BW according to congestion
  - each multicast address consumes routing table entry

Networks

- Physical Distribution Networks
  - postal service, parcel services, convenience stores
- Information Communication Networks
  - Publishing Network (Book, News Paper, CD, Movie)
  - Financial Network
  - Phone Network
  - Broadcast Network
  - the Internet

Publishing Network

- Mass Distribution of Same Information
- Delay of the Distribution may be Tolerated
- Protected by Copyright Act
- The First Victim of the Internet
  - Collapsing

Financial Network

- Manage Transfer of Money
- Partly, Phisical Distribution Network, but, today, mostly ICN
- Security!!!
  - Not that there is no accident
  - Who will pay the loss on accidents

Phone Network

- Network for Realtime Voice Transfer
  - Allocate bandwidth for voice transfer
  - Minimize (guarantee) delay for voice transfer
- Dedicated line service may be Offerred
  - but, primary service is voice transfer
- Slow and Expensive
- Was Protected as National Company
  - Leberated by Telecommunication Business Act

Broadcast Network

- Network to Transfer Voice/Image to Many in Realtime
  - Allocate bandwidth for the transfer
  - Minimize delay
- Wide Area One to Many Communication over Radio Waves
  - Broadcast/Multicast
- Protected by Broadcast Act

broadcast	phone	data comm
broadcast network	phone network	dedicated line

networks before the Internet
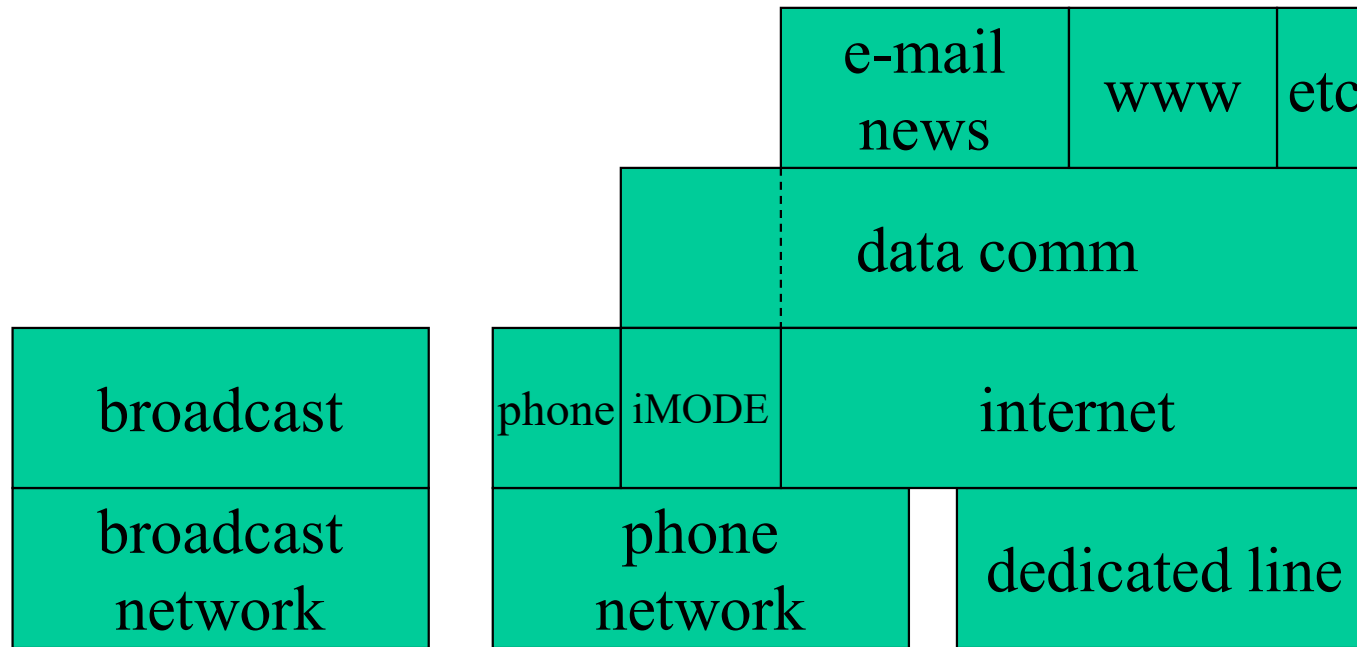
networks with the Internet

broadcast	phone	e-mail news	www	etc
streaming		data comm (batch)		
internet				
dedicated line (including wireless)				

network in the future

Multicast and Broadcast

- broadcast
  - send to all the hosts within a region
  - not realistic over the entire internet
- multicast
  - send to all the members of a group
    - # of members can be arbitrary large
      - member management by network impossible
      - members tell network their existence
  - group is identified by multicast address
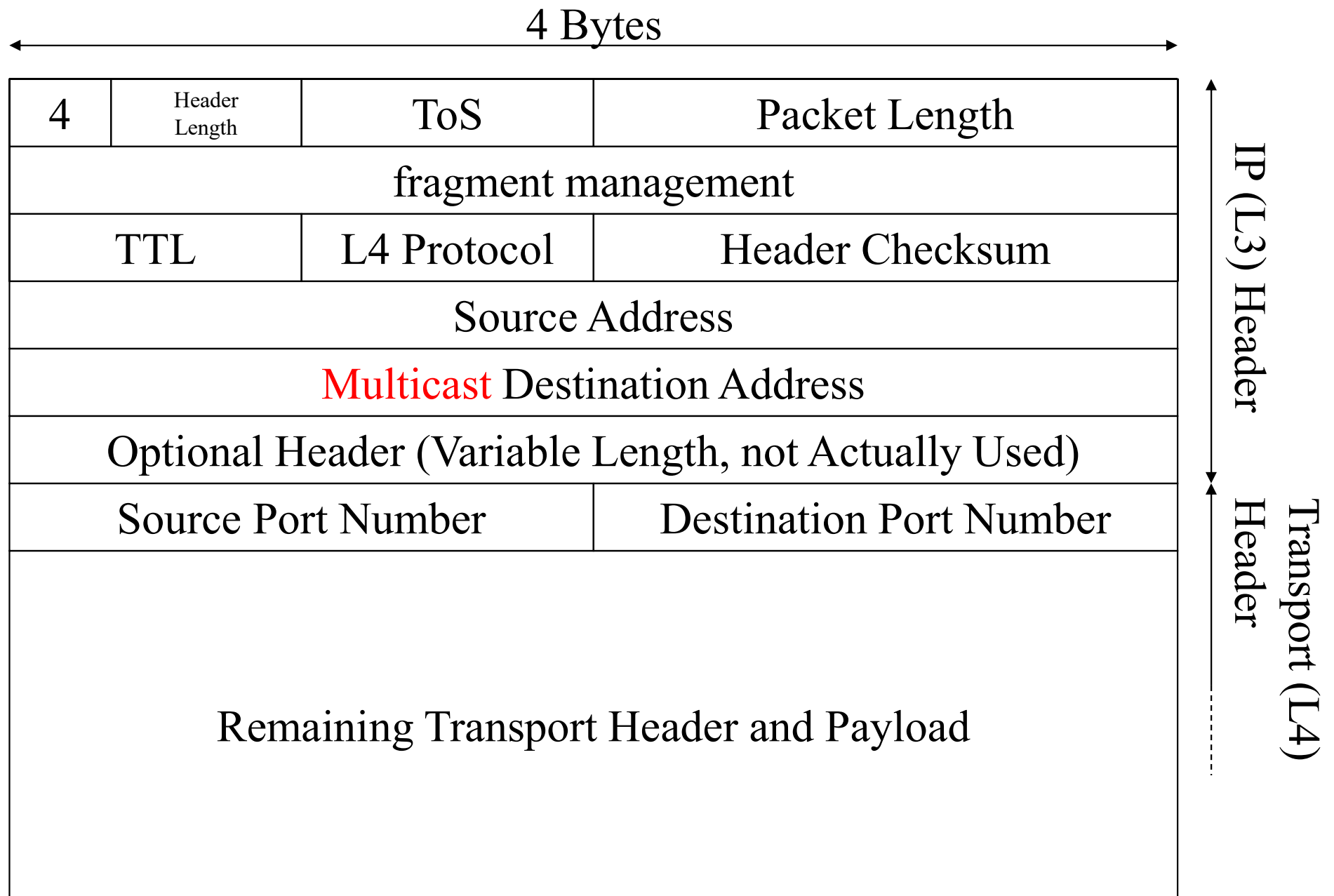   - 224.0.0.0~239.255.255.255
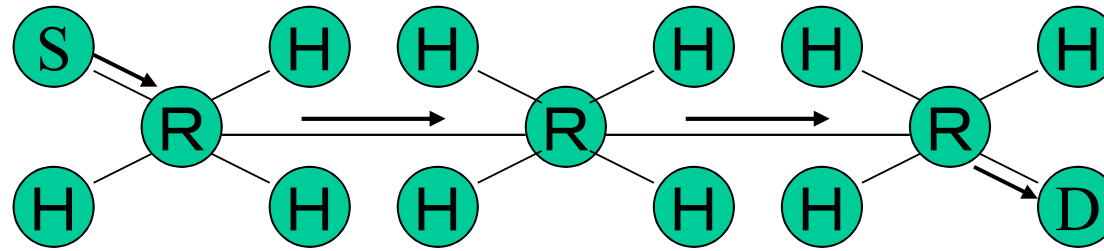
a) unicast
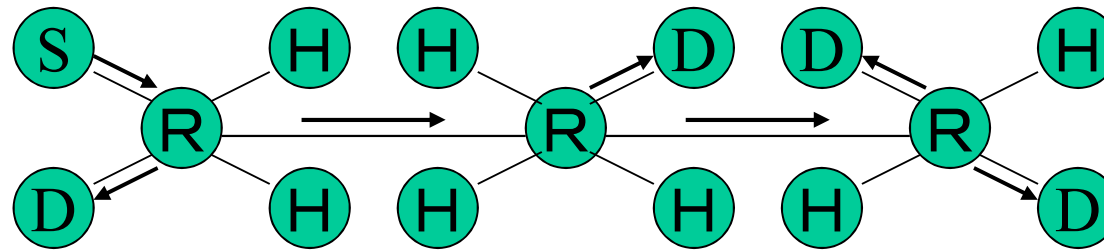
b) multicast
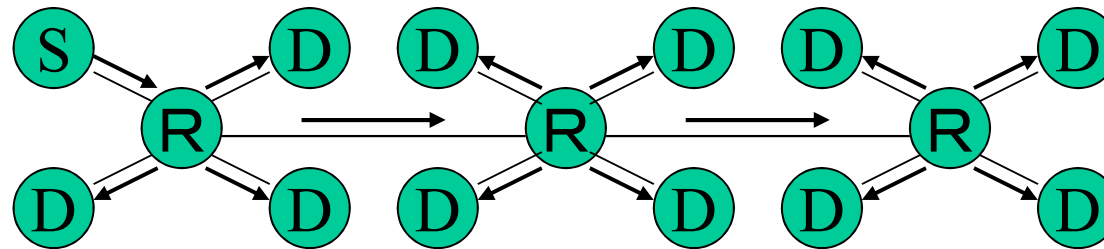
c) broadcast

(H): host $\quad$ (R): router $\quad$ (S): source $\quad$ (D): destination $\quad$ $\longrightarrow$: data flow

unicast, multicast and broadcast

4 Bytes			
4	Header Length	ToS	Packet Length
fragment management			
TTL		L4 Protocol	Header Checksum
Source Address			
Multicast Destination Address			
Optional Header (Variable Length, not Actually Used)			
Source Port Number		Destination Port Number	
Remaining Transport Header and Payload			

IP (L3) Header

Transport (L4) Header

Format of IPv4 Packets (rfc791)

Multicast by IGMP (1)

- **destination hosts**
  - changes dynamically
  - register their existence by IGMP (Internet Group Management Protocol, rfc988)
    - IGMP is independent from multicast routing protocols (?)

- **source hosts**
  - changes dynamically
  - just send multicast packets
  - independent from multicast routing protocols

Multicast by IGMP (2)

- routers
    - process some multicast routing protocol
    - depending on multicast routing protocol
        - react to IGMP packets
        - react to multicast packets sent
    - against the E2E principle?

Multicast and Ends

- destination: represented by destination host
- source: represented by source host
- group: represented by ?
  - no one?
  - ISP?
  - source (SSM, single source multicast)?
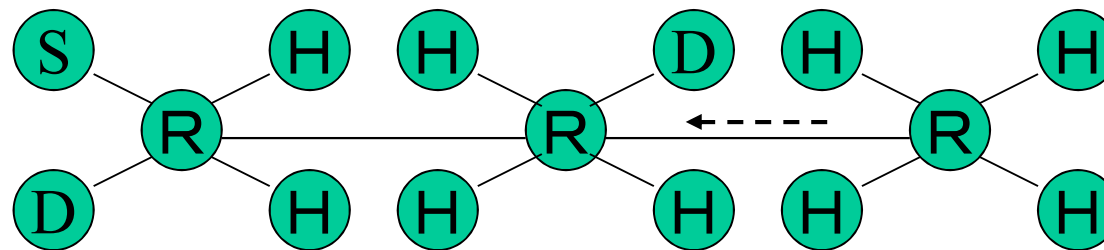 - group management host!

Multicast Routing Protocols

- dense
  - broadcast data and detect part of network where data is not necessary
  - DVMRP (rfc1075), PIM-DM
- MOSPF (rfc1584)
  - broadcast locations of sources and destinations
- sparse
  - have a center to control data flow
  - CBT (rfc2189), PIM-SM(rfc2362)
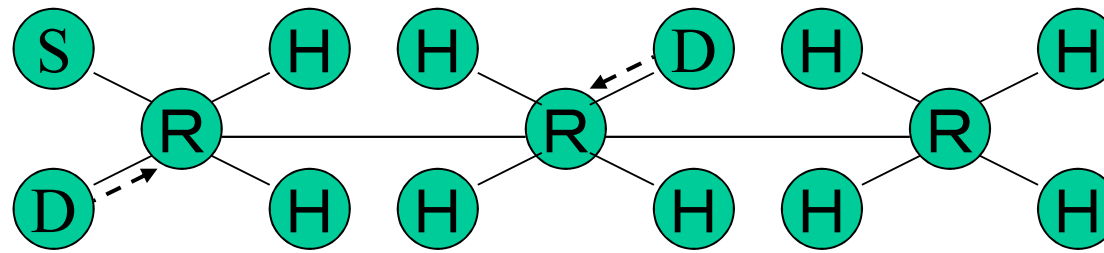
a) registration by IGMP
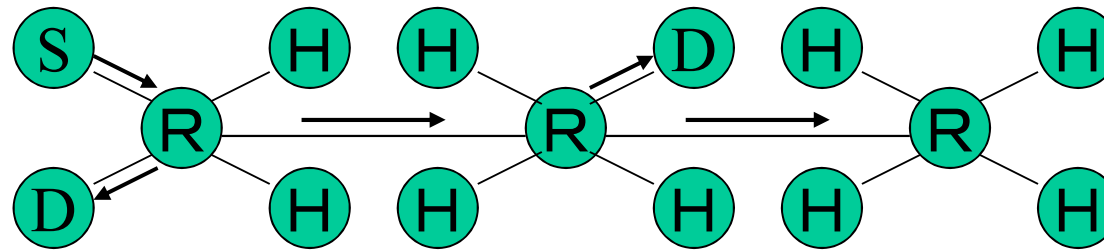
b) first data flow
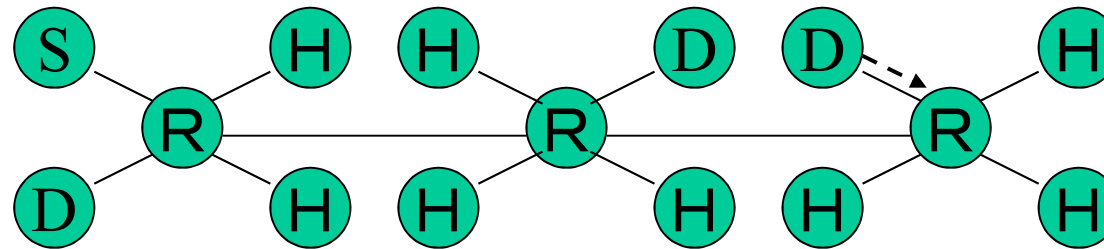
c) delete leaf with no destination

H: host   R: router   S: source   D: destination   →: data flow
                                                         --→: control flow
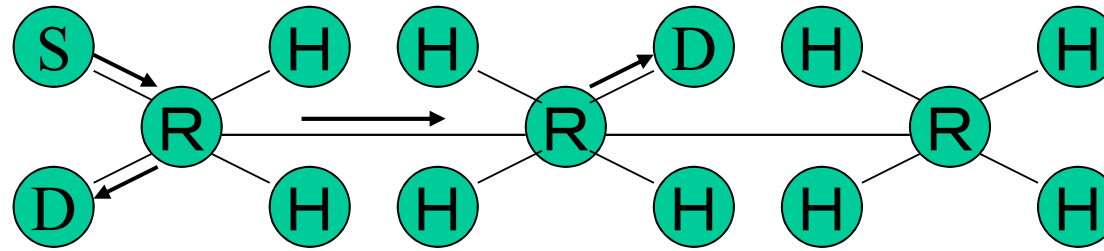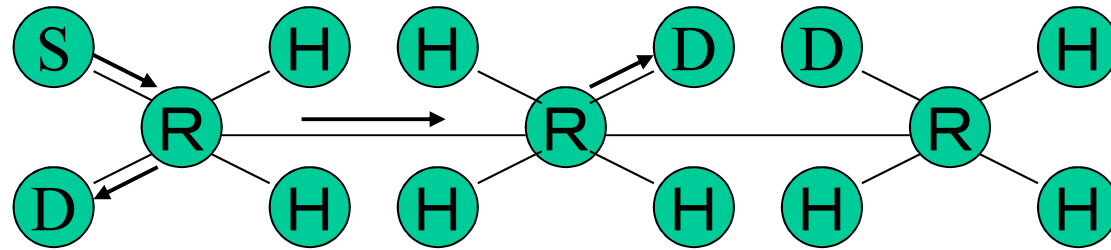
operation of DVMRP

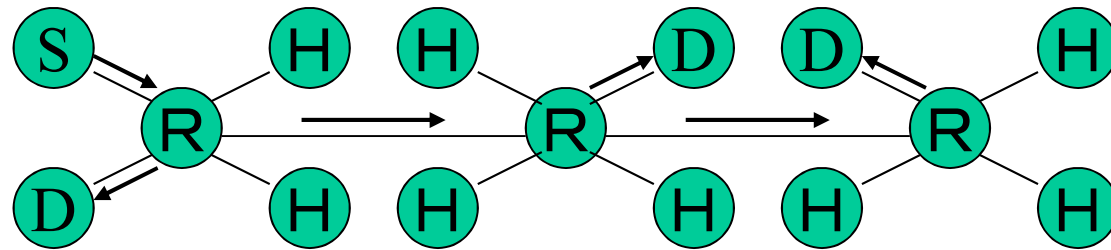d) data flow after deletion
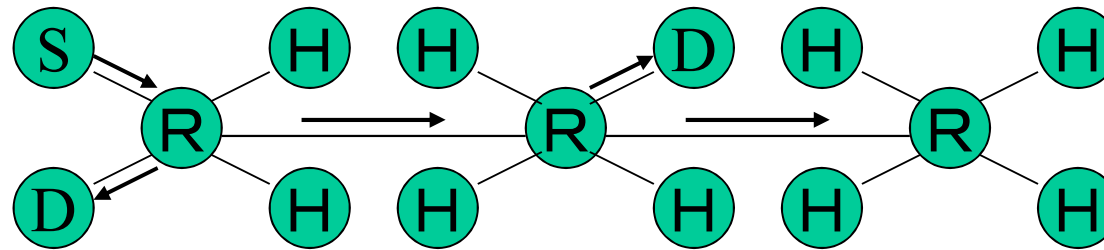
e) data flow after certain period of time

f) new destination appears

H: host   R: router   S: source   D: destination   →: data flow
                                                        --→: control flow
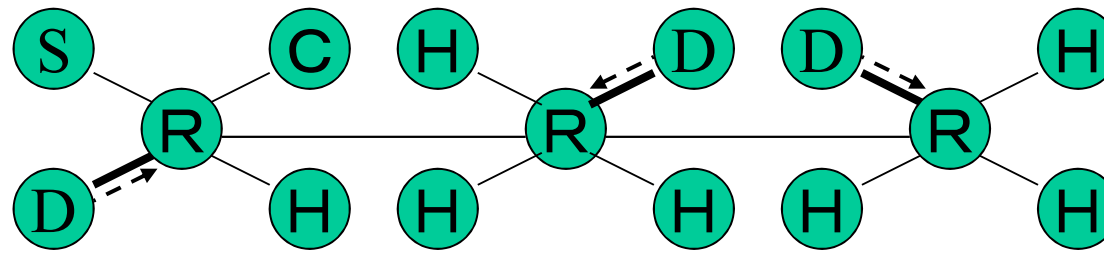operation of DVMRP

g) data flow immediately after f)

h) data flow after certain period of time

(H): host   (R): router   (S): source   (D): destination   →: data flow   --→: control flow
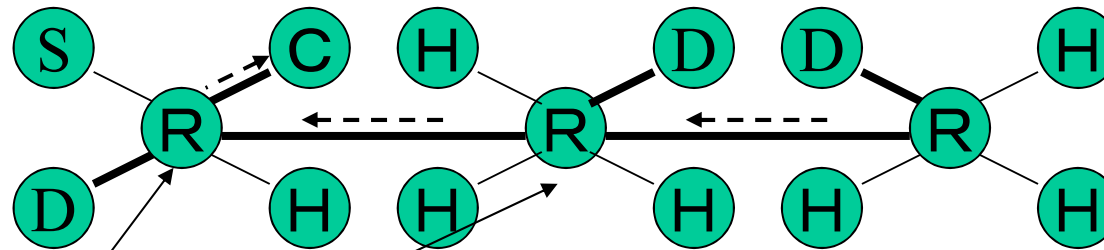
operation of DVMRP

CBT (Core Based Tree)

- routers around destinations send registration message toward Core (center)
    - multiple registration messages are merged
    - bi-directional tree including Core and destinations formed
- packets from source is relayed toward Core
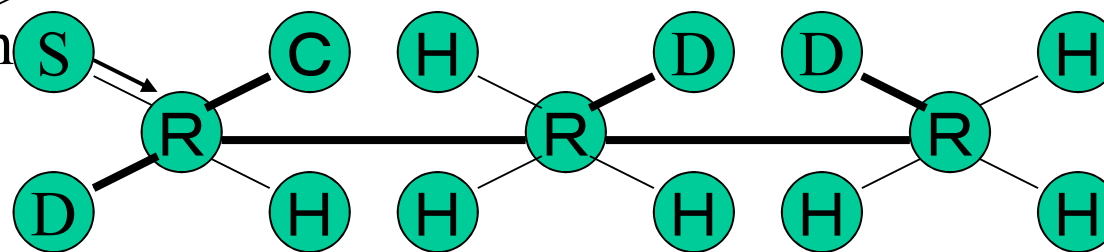   - if the packets arrives to the bi-directional tree, copied over the tree

a) registration of destinations

——: multicast tree

b) propagation of registration requests (tree is formed)

merge
registration
requests

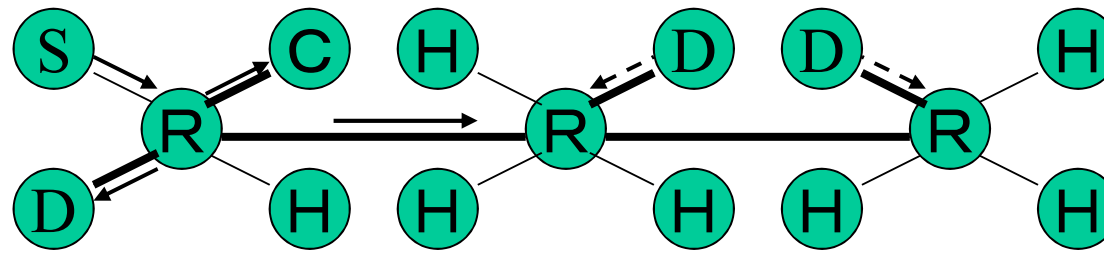c) sending packets toward core

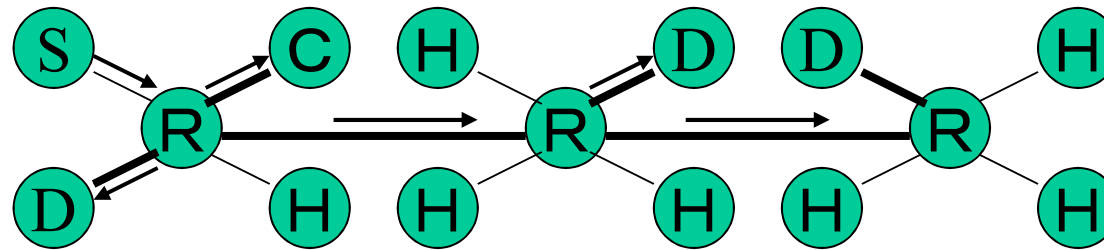H: host R: router S: source D: destination →: data flow --→: control flow
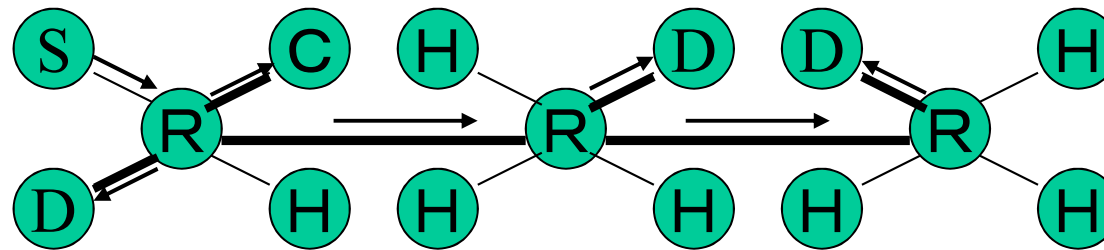
operation of CBT

d) packet copied over tree 1

—— : multicast tree

e) packet copied over tree 2

f) packet copied over tree 3

(H): host   (R): router   (S): source   (D): destination   —→ : data flow

- - -→ : control flow

operation of CBT

PIM (Protocol Independent Multicast) SM (Sparse Mode)

- routers around destinations send registration message toward RP (Rendez-vous Point)
  - multiple registration messages are merged
  - uni-directional tree rooted by RP formed
- packets from (router adjacent to) source is unicast to RP
  - packets arriving RP is copied over the tree
- RP represent group management host
  - can control dataflow
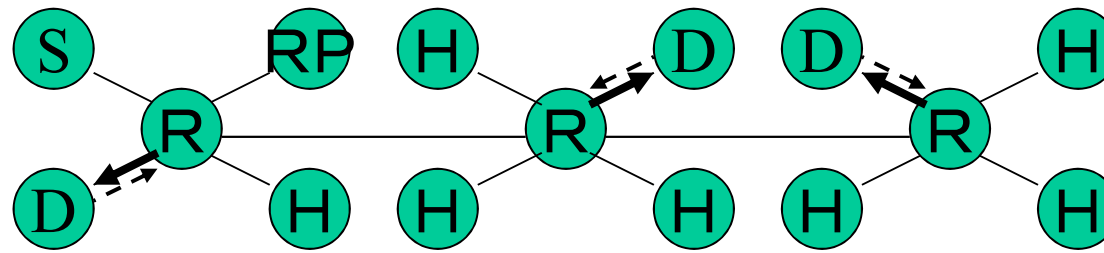
a) registration of destinations
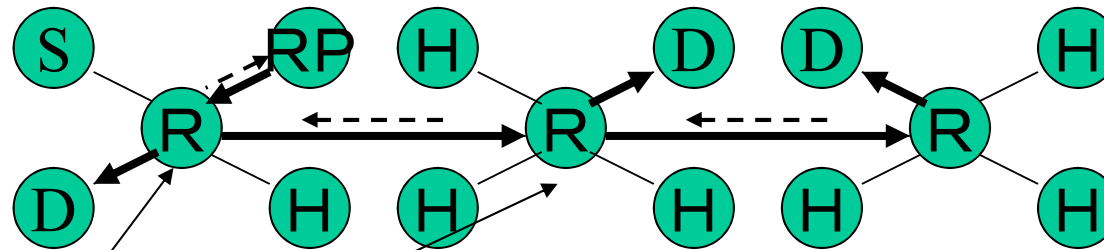
b) propagation of registration requests

$\longrightarrow$: multicast tree

merge registration requests
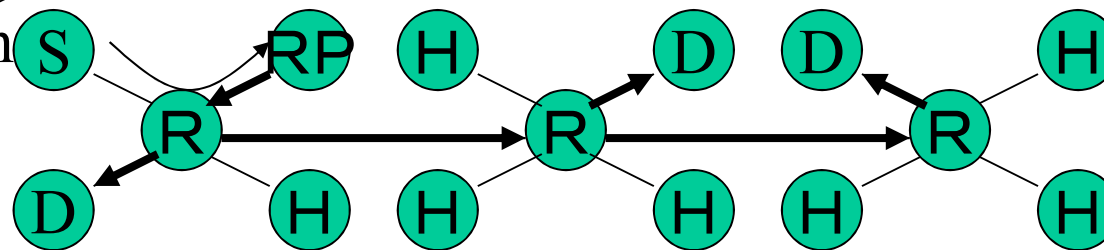
c) packets unitcast to RP (using IP tunneling)

(H): host   (R): router   (S): source   (D): destination $\longrightarrow$: data flow

$--\rightarrow$: control flow
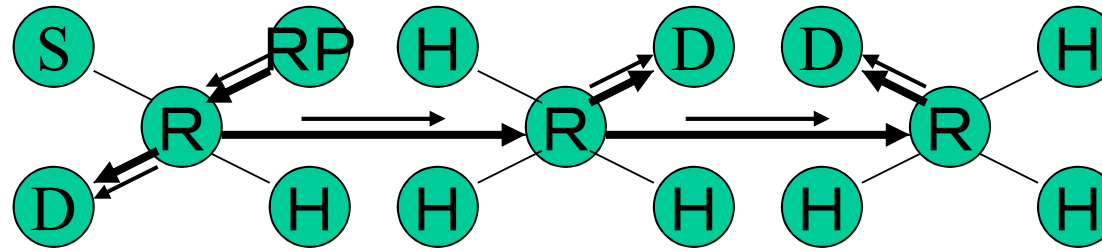
operation of PIM-SM

d) relay by RP (multicast)

: multicast tree

H : host   R : router   S : source   D : destination   : data flow

operation of PIM-SM

--→: control flow

Core and RP

- how to know Core and RP of a group?
  - broadcast?
  - delivered by application along with multicast address?
    - what if, a router receives inconsistent information?
  - give up many to many and make source RP
  - static multicast!
    - Core and RP registered to reverse DNS domain

Interdomain Multicast

- existing multicast protocols needs separate routing table entry for each group
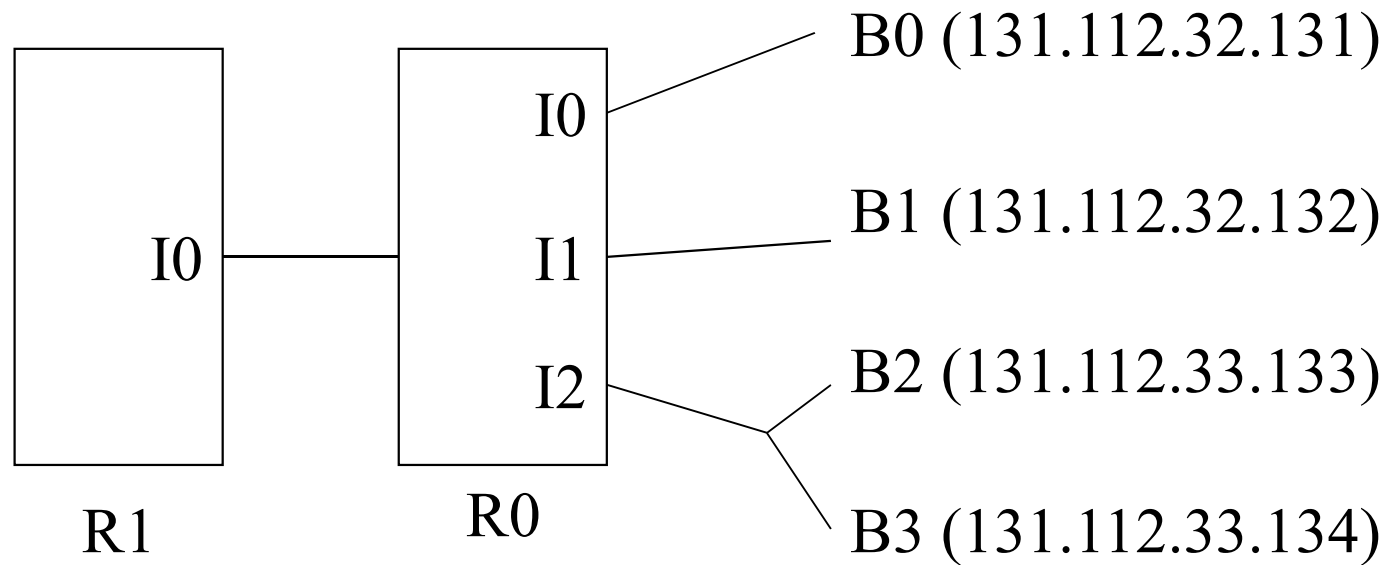 - does not scale over the Internet?
- existing protocols should be used in small domain
- provide other protocol for interdomain routing
  - BGMP (Border Gateway Multicast Protocol, rfc3913, something like Interdomain CBT)
- each domain has block of multicast address and aggregate routing table entry between domains

Aggregation of Multicast Routing Table Entries

- impossible

- multicast address does not designate location

- distribution of destinations is different group by group (even with similar address)

- if multicast has a center (center domain of BGMP)
  - route from source to the center (essentially unicast) can be aggregated
  - route from the center to destinations cannot be aggregated

Routing Table

- routers send packets to next hop routers based on look up results of routing table

  - key of the look up is destination address

- same entry may be shared if similar(?) addresses occur only in some remote region

  - route aggregation

    - 1 entry shared by many addresses

  - like phone numbers, may be hierarchical

    - +81-3-5734-3299

B0 (131.112.32.131)

B1 (131.112.32.132)

B2 (131.112.33.133)

B3 (131.112.33.134)

I0

I1

I2

I0

R1

R0

routing table at R0

destination	next hop
131.112.32.131	I0
131.112.32.132	I1
131.112.33.*	I2

routing table at R1

destination	next hop
131.112.*	I0

route aggregation

Cases When Route Aggregation Impossible

- aggregation possible, if route is shared by addresses sharing a pattern

- route not by destination address only
  - QoS routing depends on required QoS

- destination address not designate location
  - multicast address designate set of locations

- random IP addresses within a region
  - initial allocations for IPv4
  - multihoming by routing

Multicast Routing Table Entries Cannot be Aggregated

- interdomain multicast by BGMP is illusion
  - static CBT and PIM works Interdomain
- routing table of internet backbone is large
- multicast group is resource reserving communication occupying limited resource of routing table entries
  - should be charged proportional to duration of the communication

Multicast and Bandwidth

- congestion situation is different by each destination

  – BW management for which destination?

- source determines BW

  – destinations somehow (BW (QoS) guarantee?) receives or give up

Economic Incentive for Multicast

- with flat rate best effort
  - ISP want to collect extra money for multicast
    - destinations have no merit to use multicast
      - easy to insist on unicast
    - source do not want to use multicast with no destinations
- with proportional charge (with QoS)
  - ISP prefer unicast
    - source/destinations want to reduce charge by multicast

Example of Multicast Cost
(8k broadcast in a prefecture)

- assume 10Gbps prefecture backbone (excluding access)
  - ISP charge 5000Yen/month, 2.5 person for each subscriber, 30% for prefecture backbone
    - backbone cost: 600Yen/(person・month)
    - occupying 100Mbps costs 72Yen/(person・year)
      - 72MYen/year with 1M populations

- multicast is
  - less expensive than radio-wave broadcast or CDN
  - costs 1/10 if backbone is 100Gbps

Wrap-up

- multicast is function of network
  - impossible by ends
- IGMP is against E2E principle
- end systems managing group is essential
- broadcast must be avoided in multicast routing protocols
- multicast route cannot be aggregated
- multicast congestion control impossible
  - multicast is resource reserving