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

8. Routing: IGP, Policy Multihoming, Mobility

Masataka Ohta

mohta@necom830.hpcl.titech.ac.jp

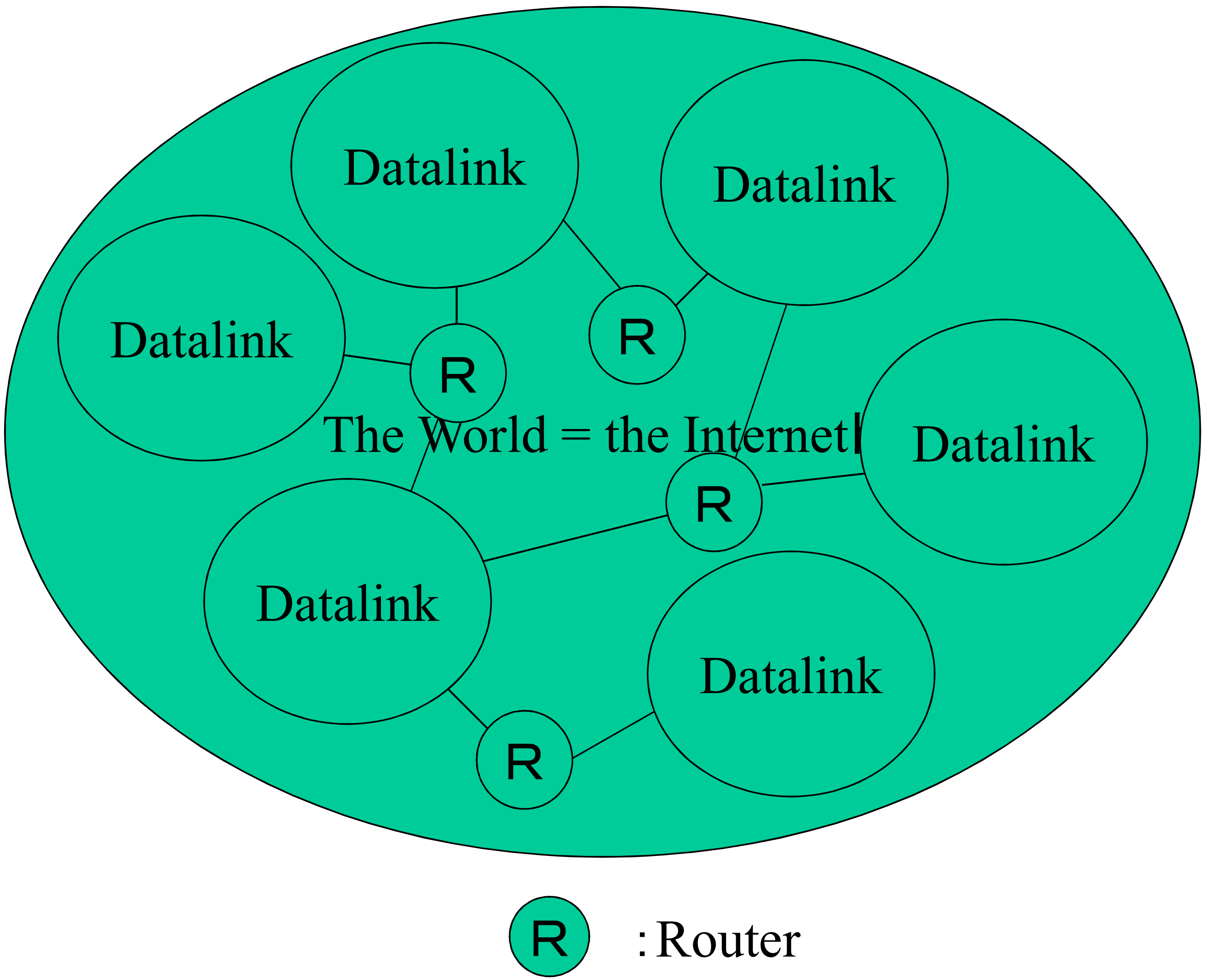
ftp://chacha.hpcl.titech.ac.jp/infra8e.ppt

What is Routing?

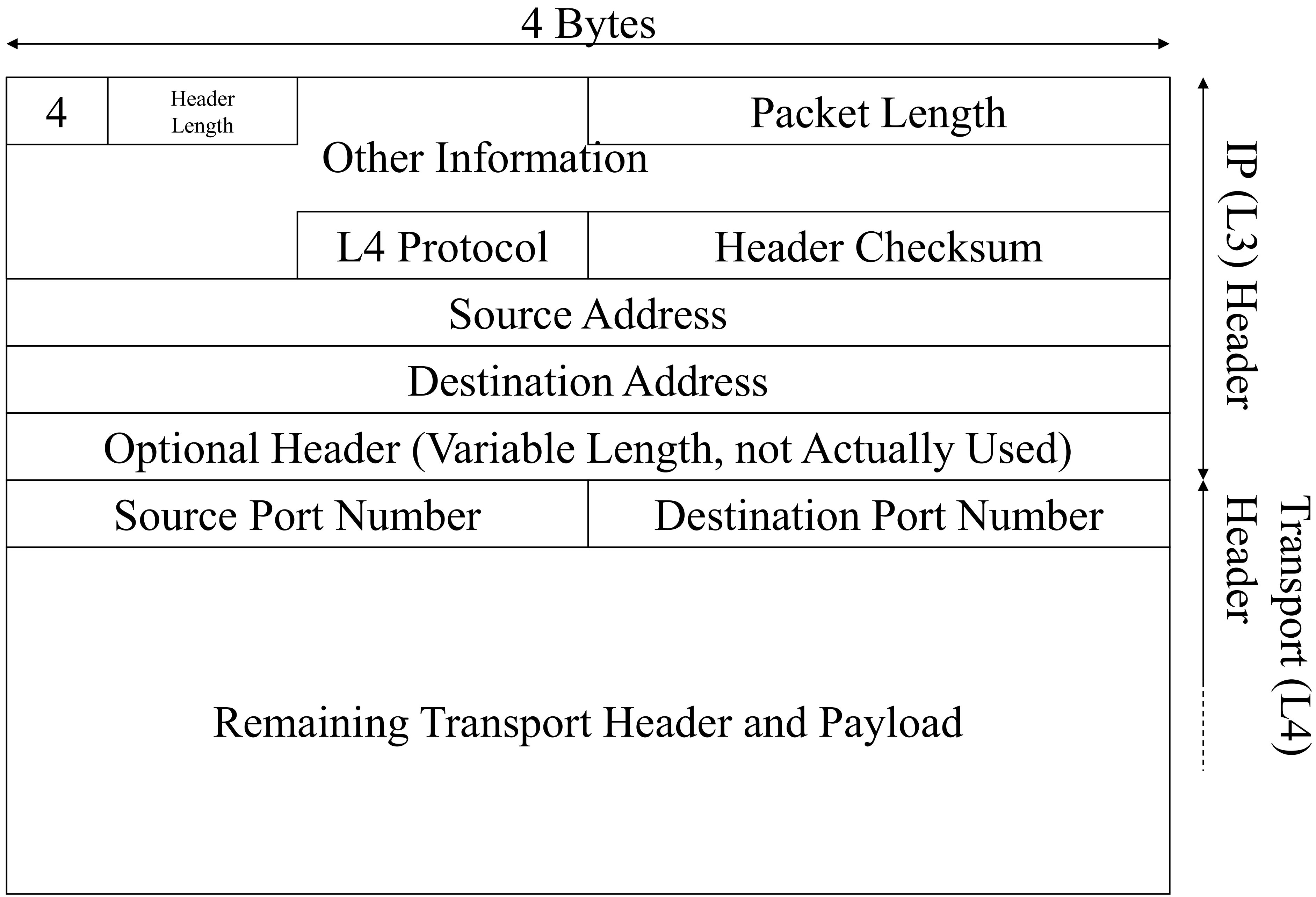
- to relay packets following routing table
  - routing table is automatically generated by routing protocols
  - defects are automatically avoided
- intelligence of the network?
  - against E2E principle?
    - the principle assumes packets are properly relayed
  - should depend on network minimally

Structure of the Internet

- CATENET Model
  - Many small (w.r.t. # of devices) datalinks interconnected by IP (Internet Protocol) routers



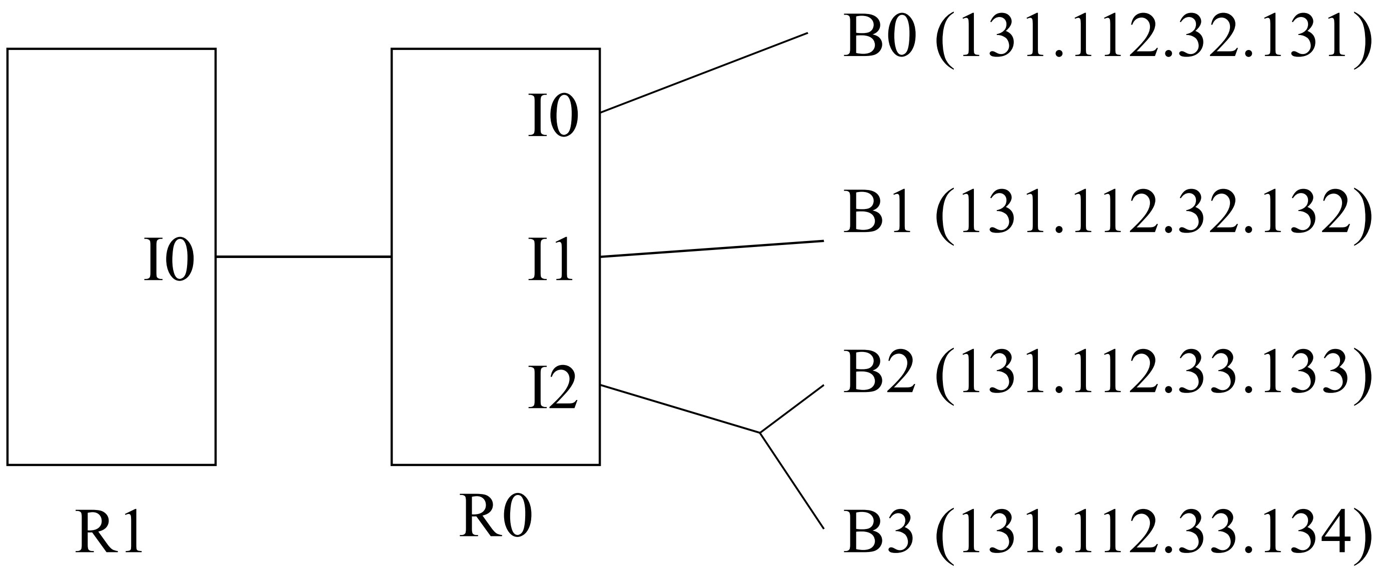
CATENET Model



Format of IPv4 Packets

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



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

Why Internet is Usually Flat Rated?

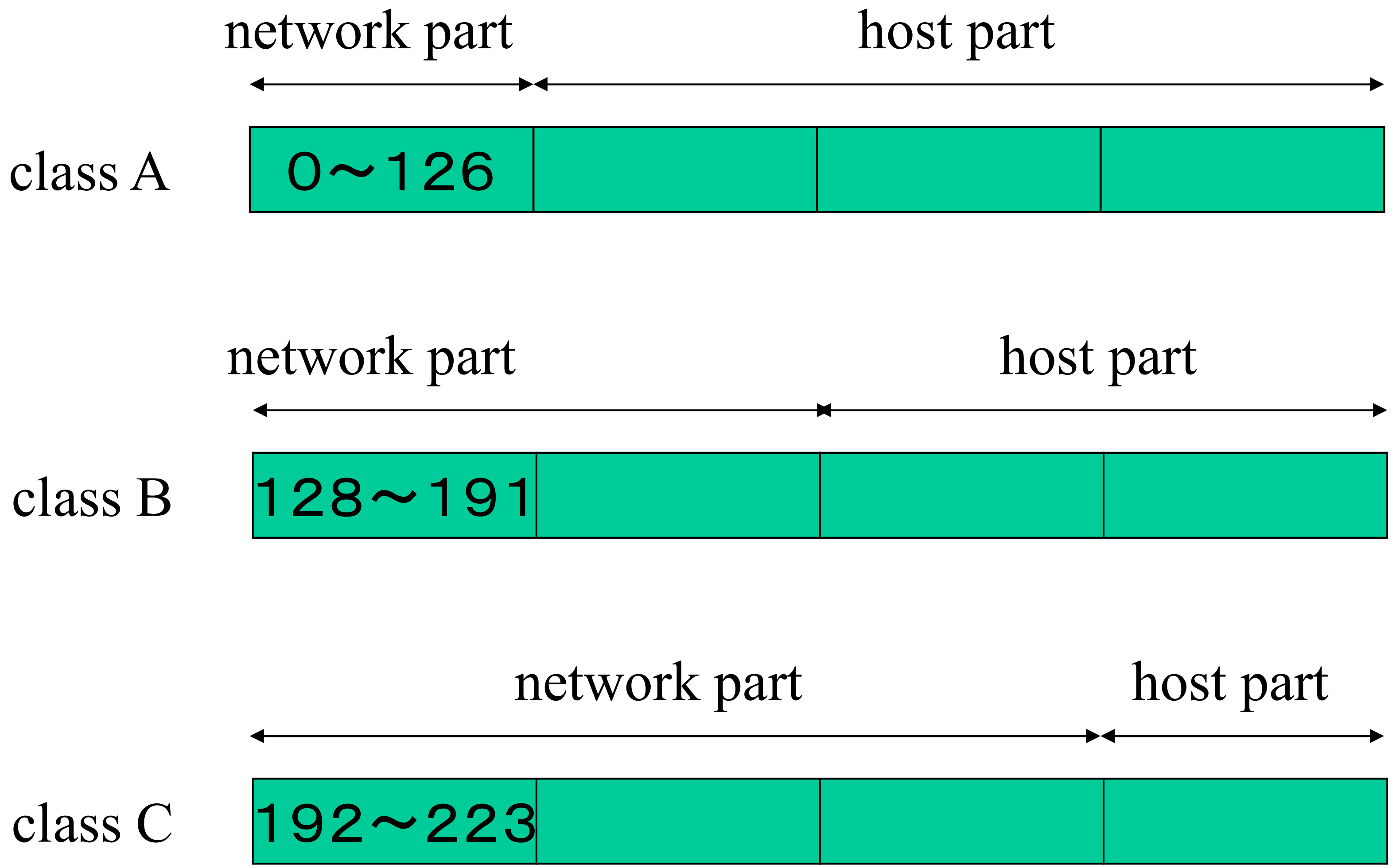
- because backbone has enough high speed?
  - was flat rated even when backbone was slow
    - is still slow compared to access
- because usage based charging costs?
  - maybe, but mobile phone limit usage
- because there is no QoS guarantee?
  - though phone quality is not very high

Internet is Flat Rated becaue Resource is not Occupied

- no prior setting of communication channel
- processing independent packet-wise
  - each packet has dst address
  - routers look up routing table by dst addresses
  - routing table entry is not occupied by each communication
    - routing table entry is finite resource
    - not necessary for each destination (multihoming?)
- Usage based charge necessary for communication occupying resource

Class based Routing

- IPv4 addresses are divided into 5 classes
  - Class A, B and C for unicast
    - class D for multicast, E reserved
- unicast IP address is divided to network part and host part
  - routing is by network part (no hierarchy)
  - host part of all 1 means broadcast within the network
  - host part of all 0 is address of the network



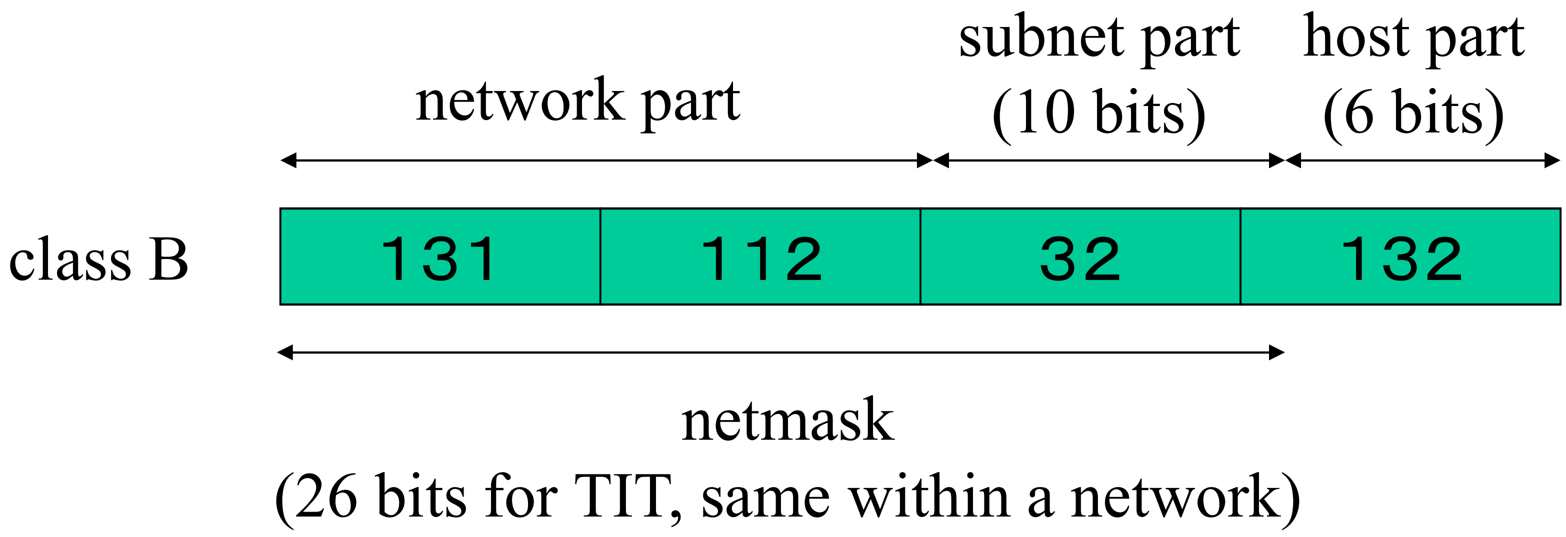
address structure of IPv4 unicast classes

Problems of Class based Routing

- each link has, at most, several tens of hosts
  - though some operated with thousands of hosts
    - only to find it inoperational
  - even class C is too large
    - unnecessary increase of route information
    - unnecessary consumption of IPv4 addresses
- finer subdivision of IPv4 address necessary
  - subnet

Subnet

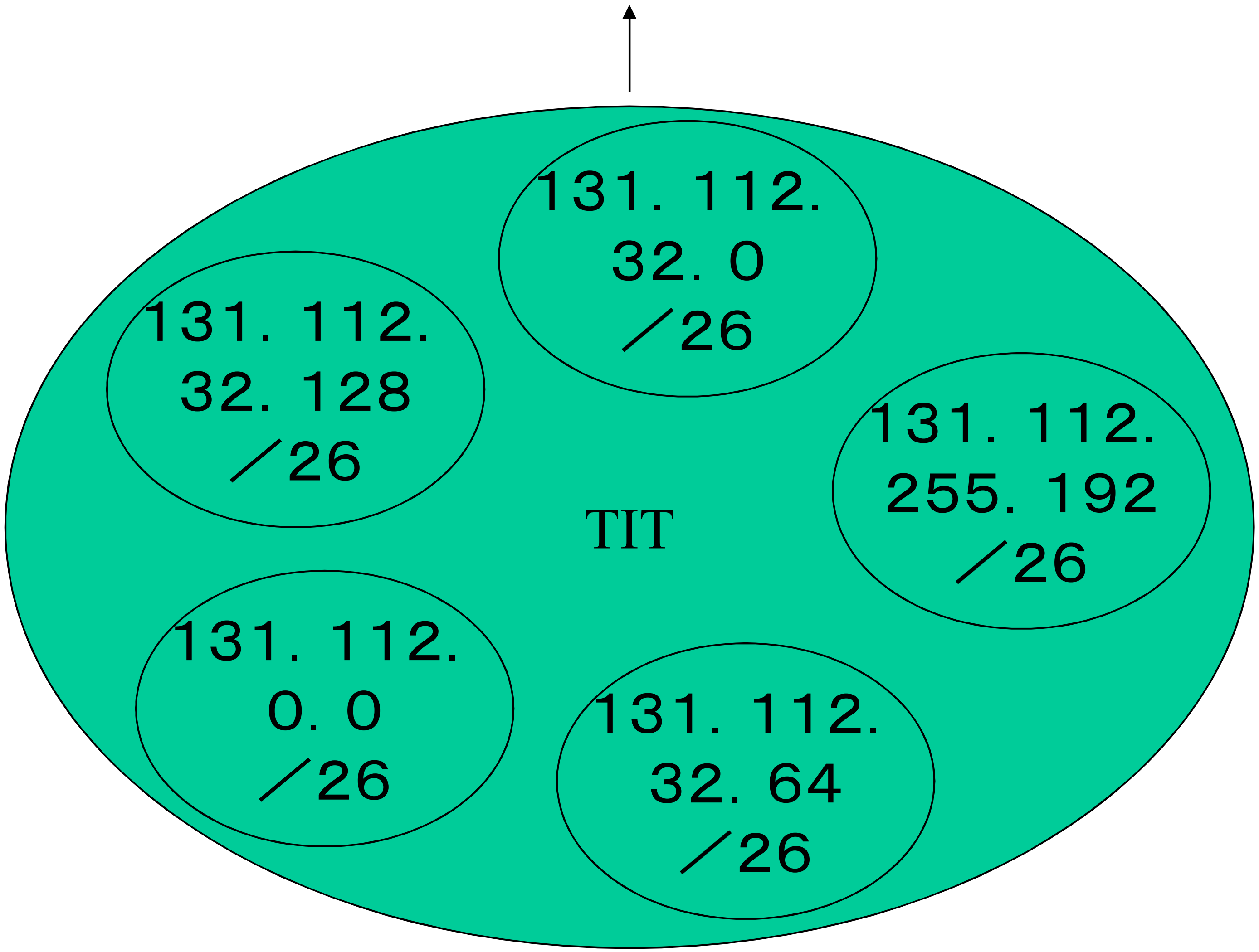
- divide host part into subnet part and host part
- subnet-wise routing within a network
  - 1 class B address is mostly enough for each organization
- network-wise routing outside of the network
  - only 1 route information for each organization externally



131.112.32.128/26 identify a subnet

example of structure of subnetted IP address of TIT

１３１．１１２．０．０／１６



１３１．１１２．３２．０／２６
１３１．１１２．３２．１２８／２６
TIT
１３１．１１２．２５５．１９２／２６
１３１．１１２．０．０／２６
１３１．１１２．３２．６４／２６

CIDR (Classless Inter-Domain Routing) (rfc1519)

- classes are totally abandoned
  - routing protocols carry netmask for each routing table entry
    - must upgrade routing protocols
- example of hierarchical address allocation
  - ISP is allocated a block of 256 addresses
    - routing table entry with netmask /24 outside of the ISP
  - the ISP allocate 8 addresses to each customer
    - 32 routing table entries with netmask /29 in the ISP for the block

Initial Proposal (rfc2374) of IPv6 Address Structure

- have strong hierarchy
- two layers at ISP level
  - TLA (Top Level Aggregater)
  - NLA (Next Level Aggregater)
- Subscribers can have 65536 links (subnets)
  - SLA (Subscriber Level Aggregater)
- 64 bit Interface ID within each link

```
| 3|  13 | 8 |   24   |   16   |          64 bits                |
+--+-----+---+--------+--------+---------------------------------+
|FP| TLA |RES|  NLA   |  SLA   |         Interface ID            |
|  | ID  |   |  ID    |  ID    |                                 |
+--+-----+---+--------+--------+---------------------------------+

<--Public Topology--->   Site
                       <-------->
                        Topology
                                <------Interface Identifier----->
```

Structure of IPv6 address

Routing Protocol

- generate routing table automatically
- two protocol styles: DV (Distance Vector) and LS (Link State)
- two environment: IGP (Interior Gateway Protocol) and EGP (External Gateway Protocol)
- RIP (rfc1056), OSPF (rfc2328), BGP (rfc1771),,,

Distance Vector Protocols

- router generate route information with metric of networks (subnets) adjacent to itself
- router receiving route information relays to other routers increasing metric
  - router receiving multiple route information to same network choose one with smaller metric
    - distributed computation of distance

Properties of DV

- small amount of computation on each router
  - because computation is distributed
- slow to react route changes
  - especially with loops

Link State Protocols

- router generate connectivity information of networks (subnets) adjacent to itself and other networks
- connectivity information is flooded to all the routers as is
- each router individually compute route with minimum distance
  - E2E principle (routers are end of routing protocols)

IGP

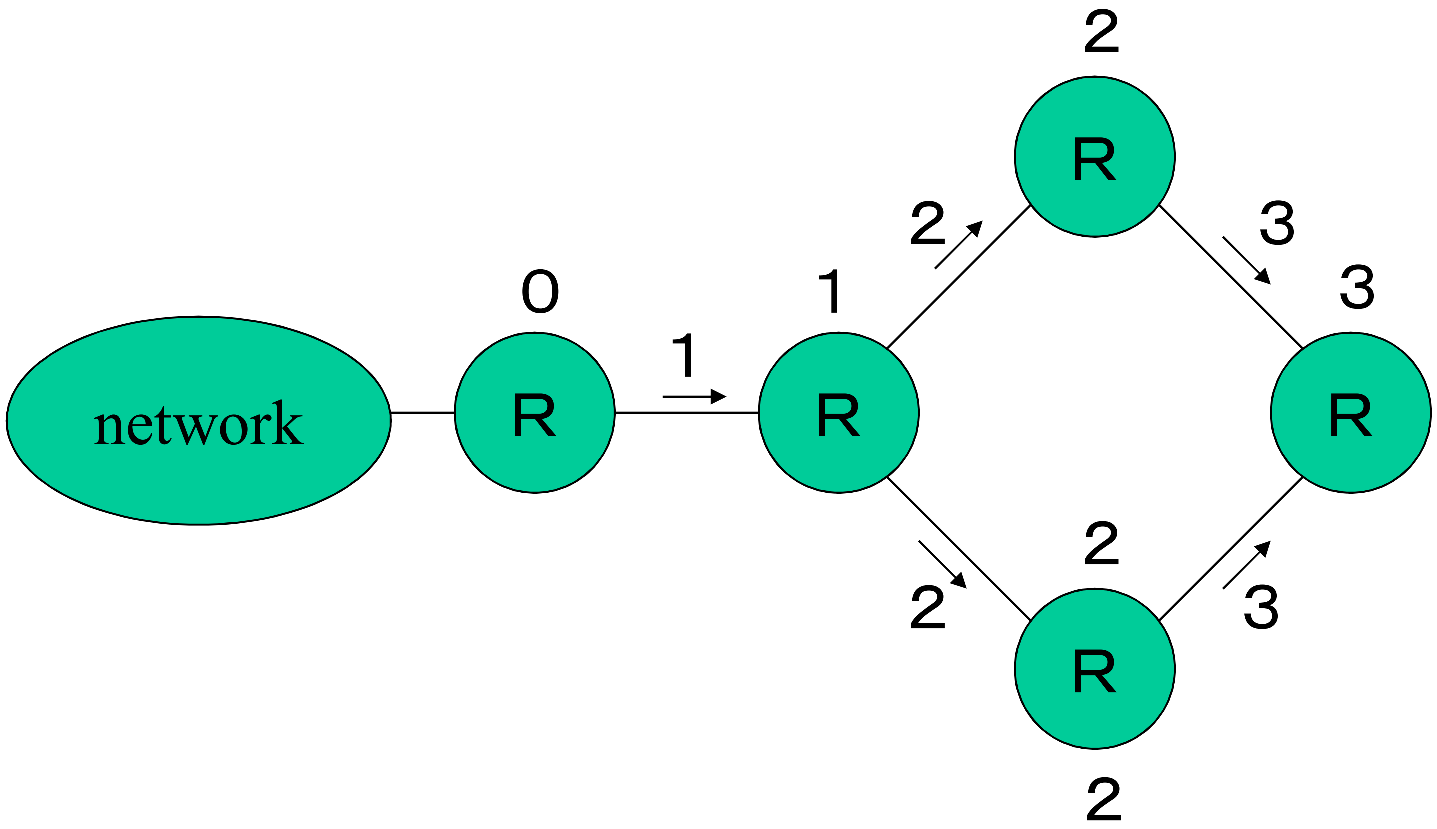
- used in a domain where everyone is cooperating
  - routing for intra site or intra ISP
- willingly carry traffic of anyone
- should choose shortest (best) path

EGP

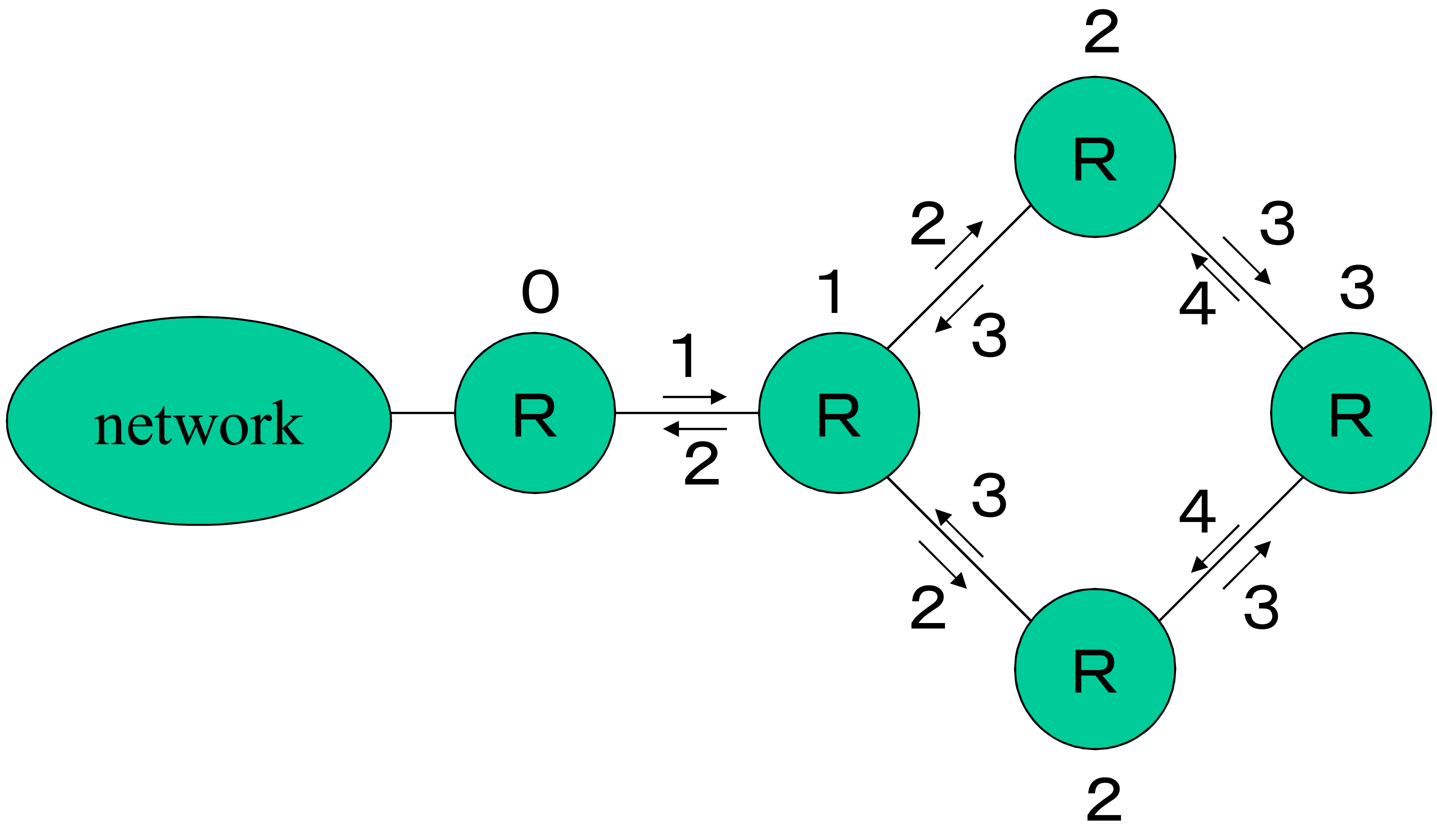
- used for routing between domains with conflicting interest
- carry traffic of others depending on negotiation (barter, money etc.)
- route is selected by policy
  - though fine control is painful

RIP (Routing Information Protocol)

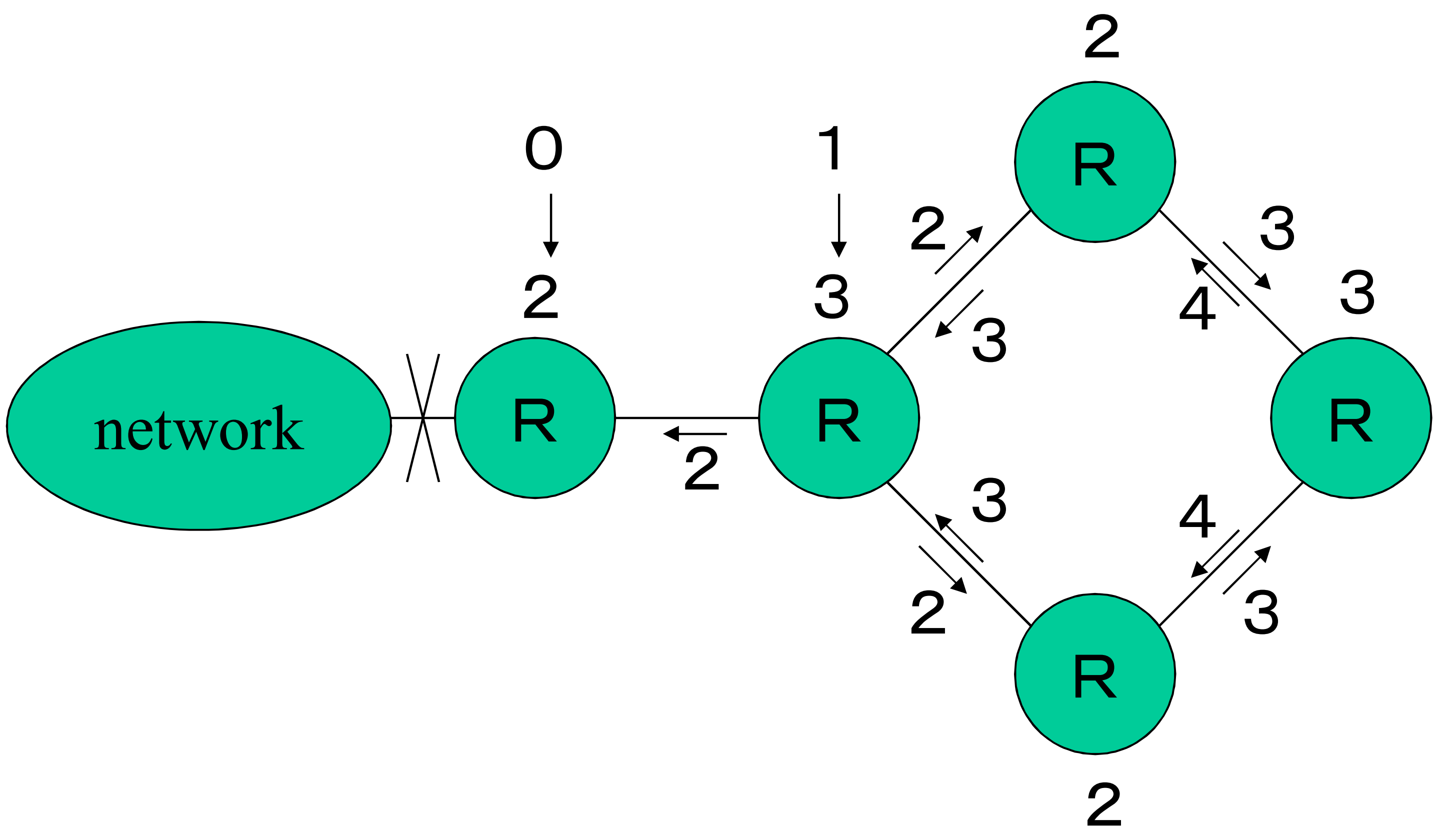
- DV style IGP
- very old
  - not CIDR capable
  - RIPv2 (rfc2453) is for CIDR
- metric is integer (0~15) (15 is infinity)
  - 15 hop was enough for the entire Internet at that time



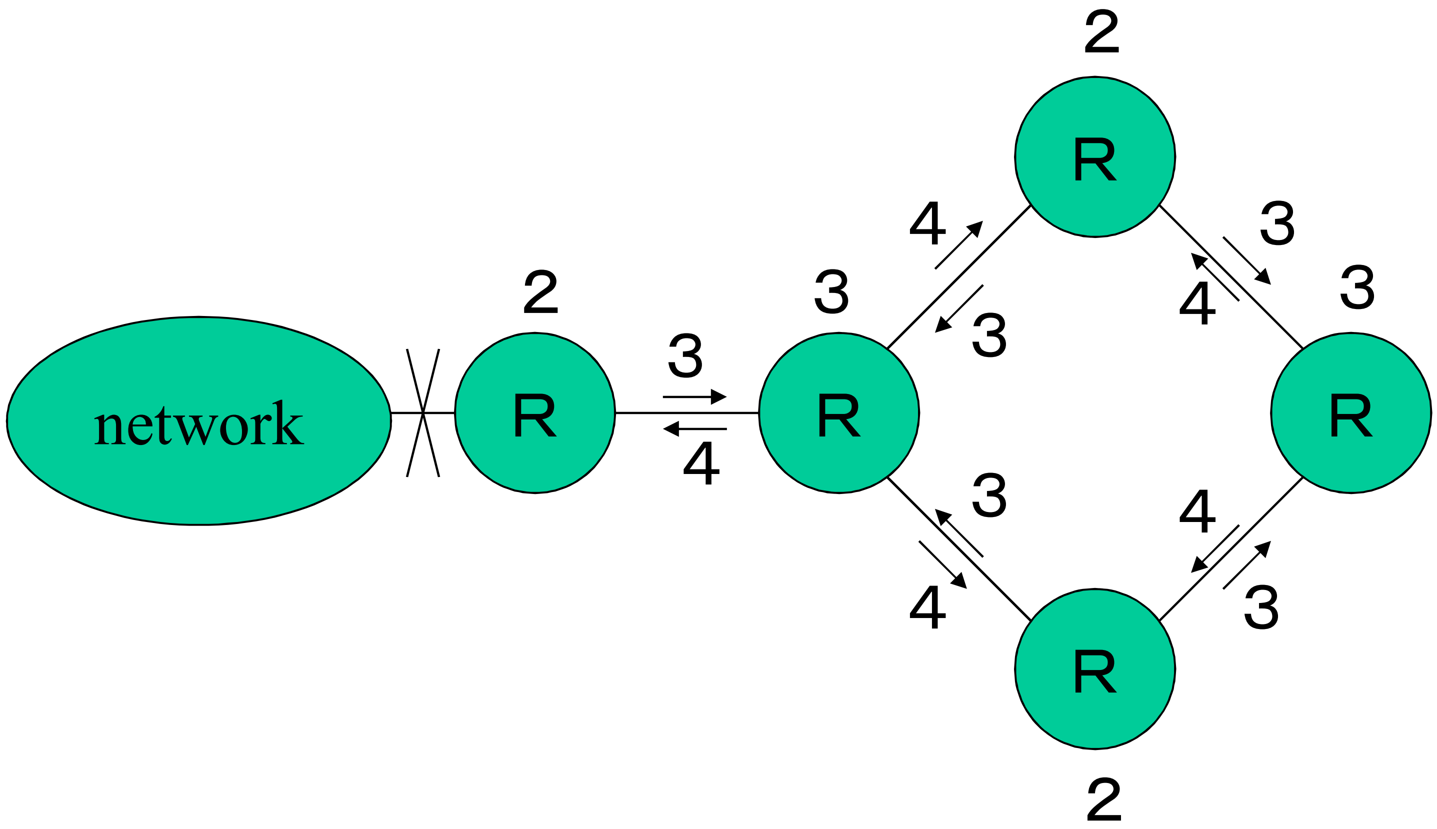
propagation of metric by RIP



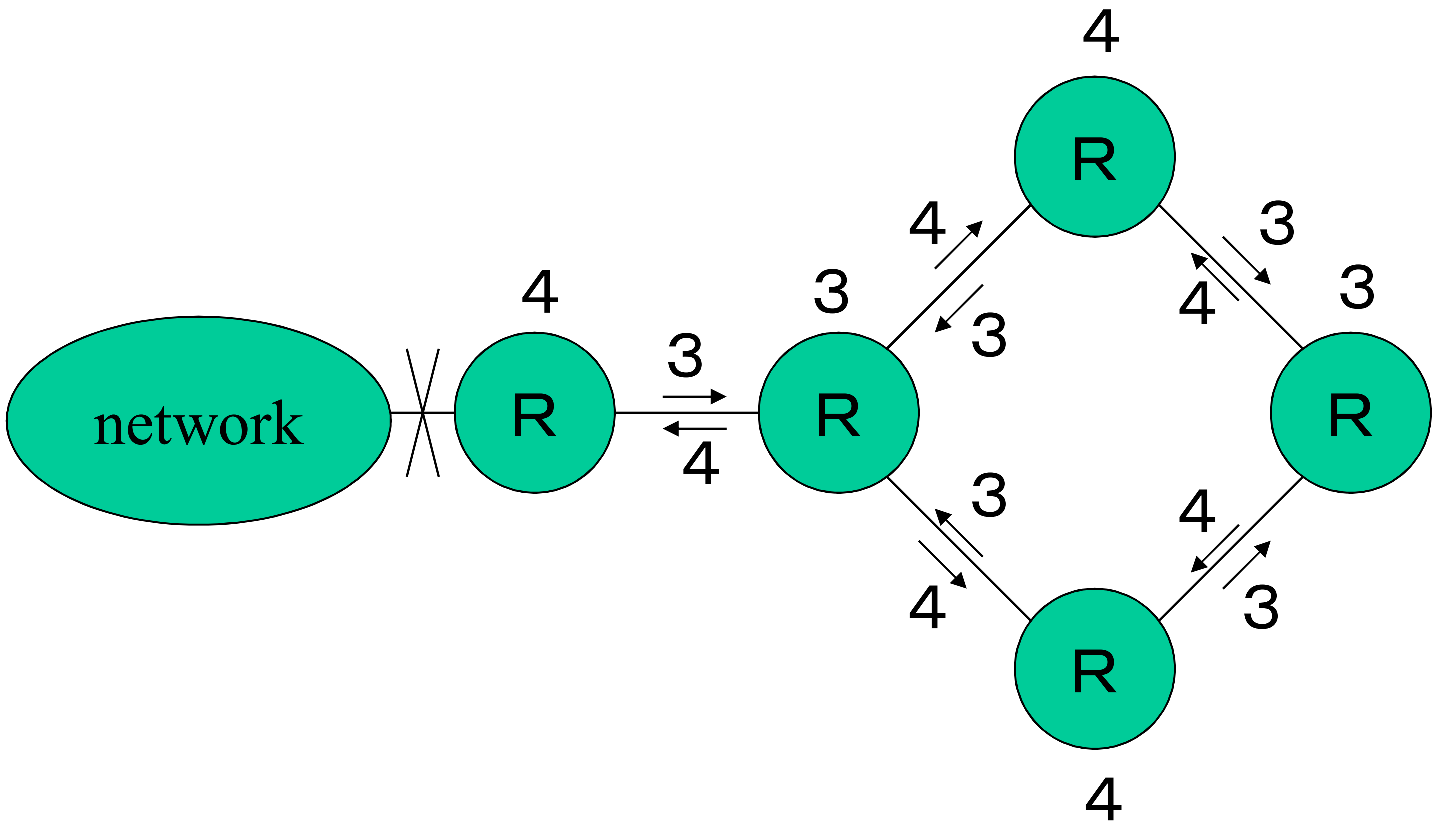
propagation of metric by RIP



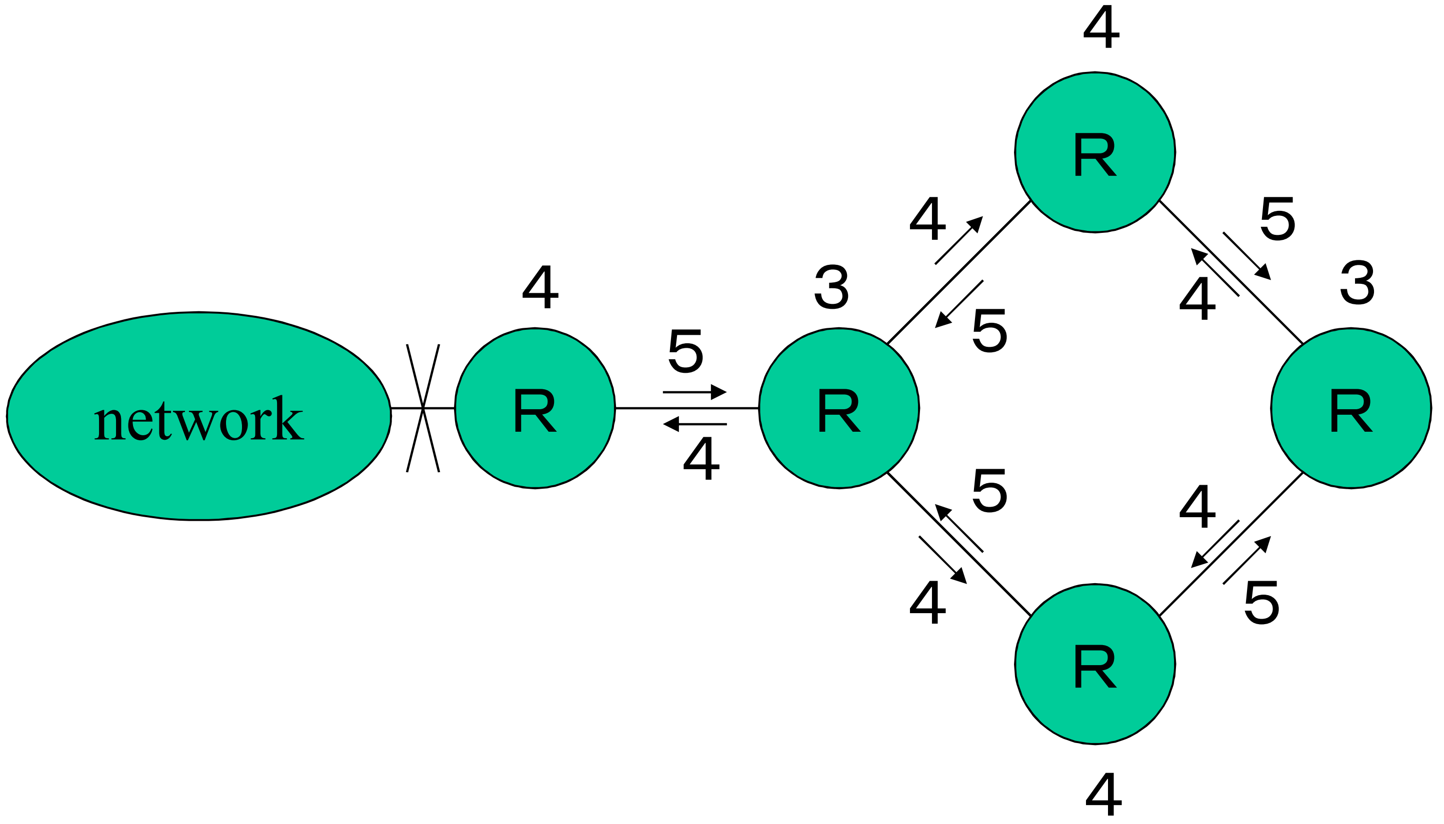
propagation of metric by RIP



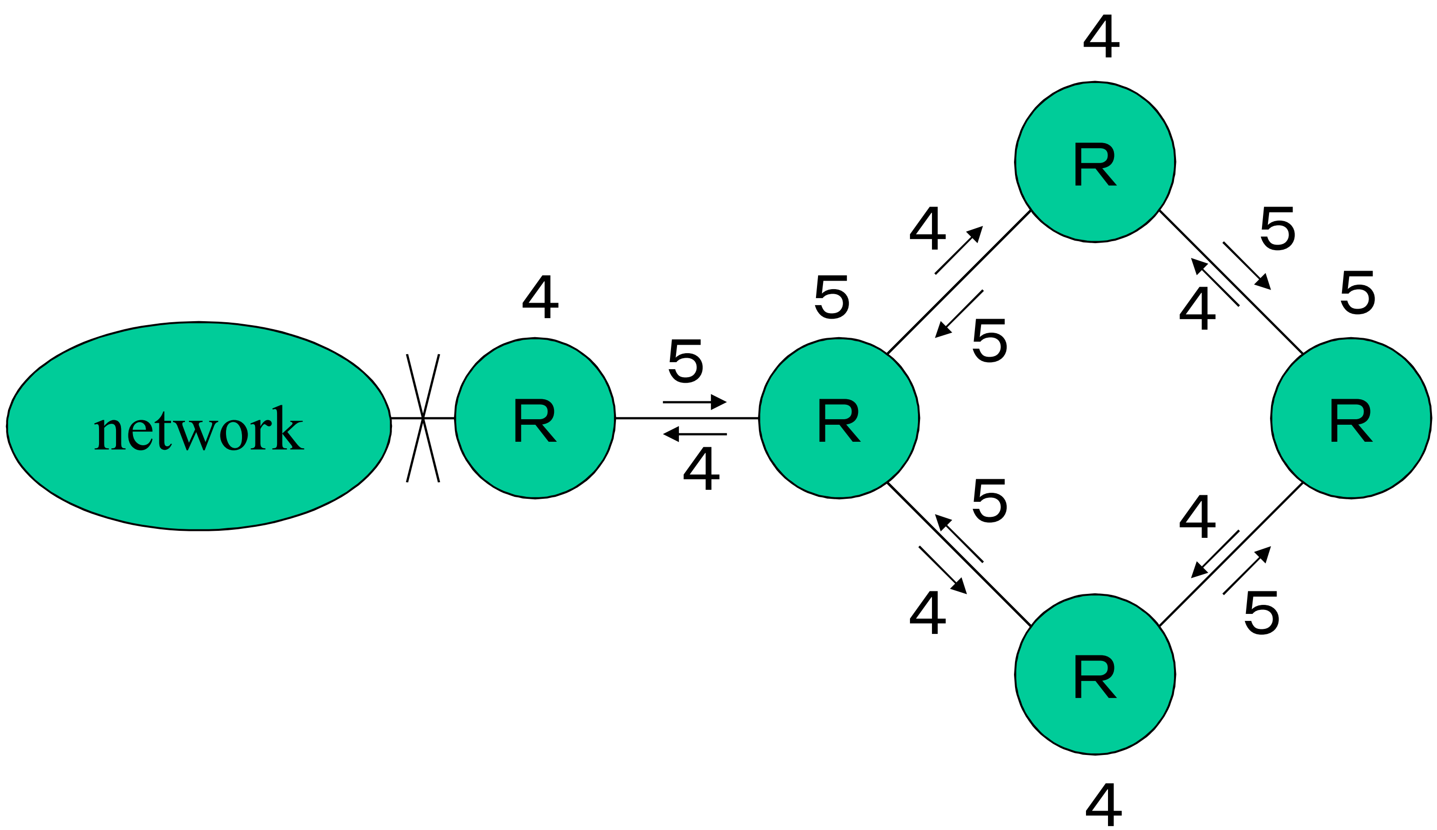
propagation of metric by RIP



propagation of metric by RIP



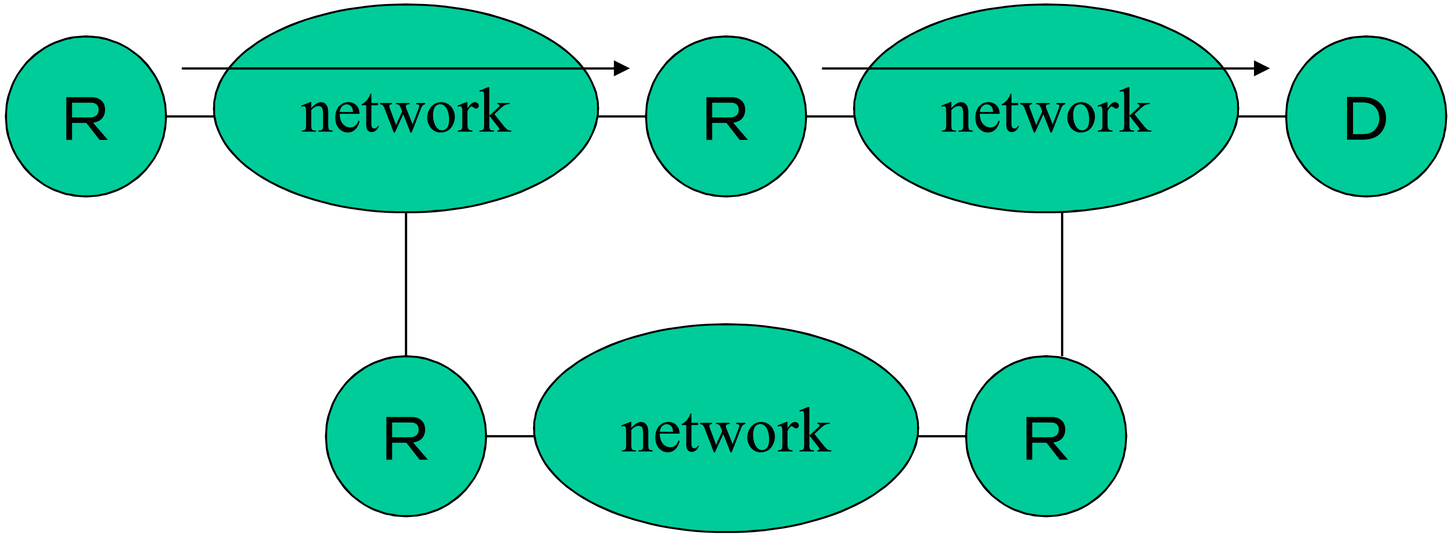
propagation of metric by RIP



propagation of metric by RIP

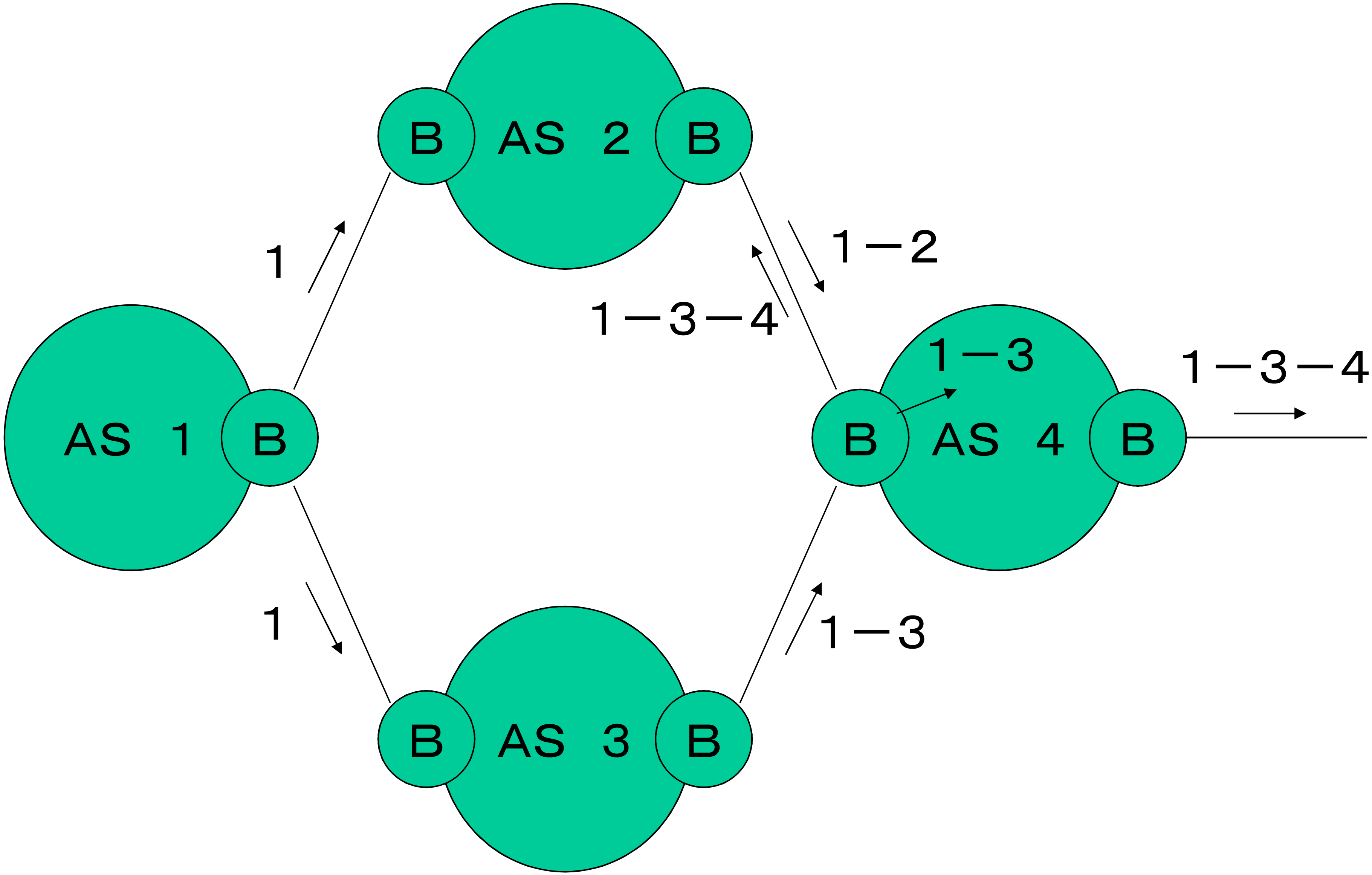
OSPF (Open Shortest Path First)

- LS style IGP
- all routers compute shortest path to all destinations based on same information

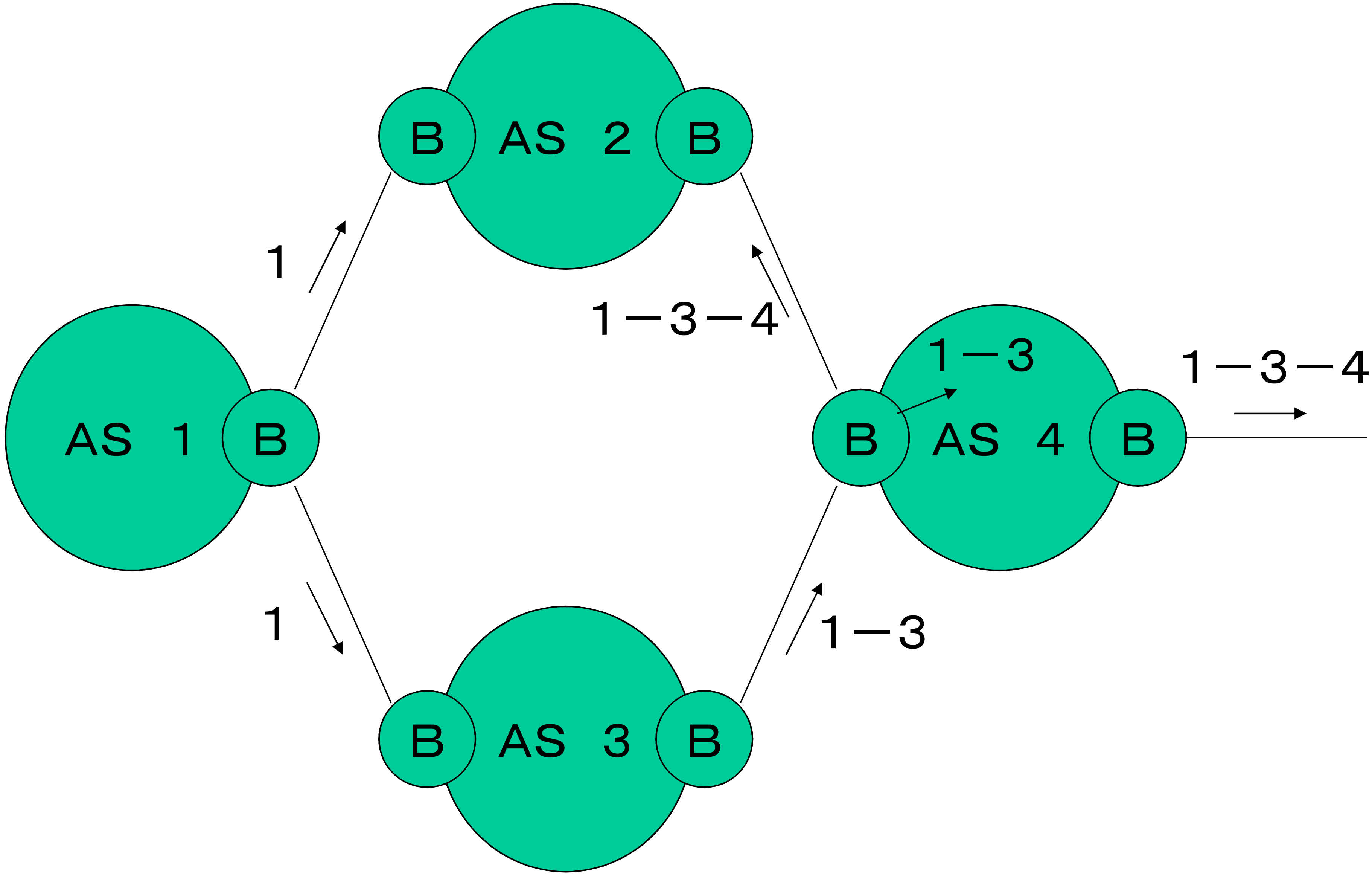


BGP (Border Gateway Protocol)

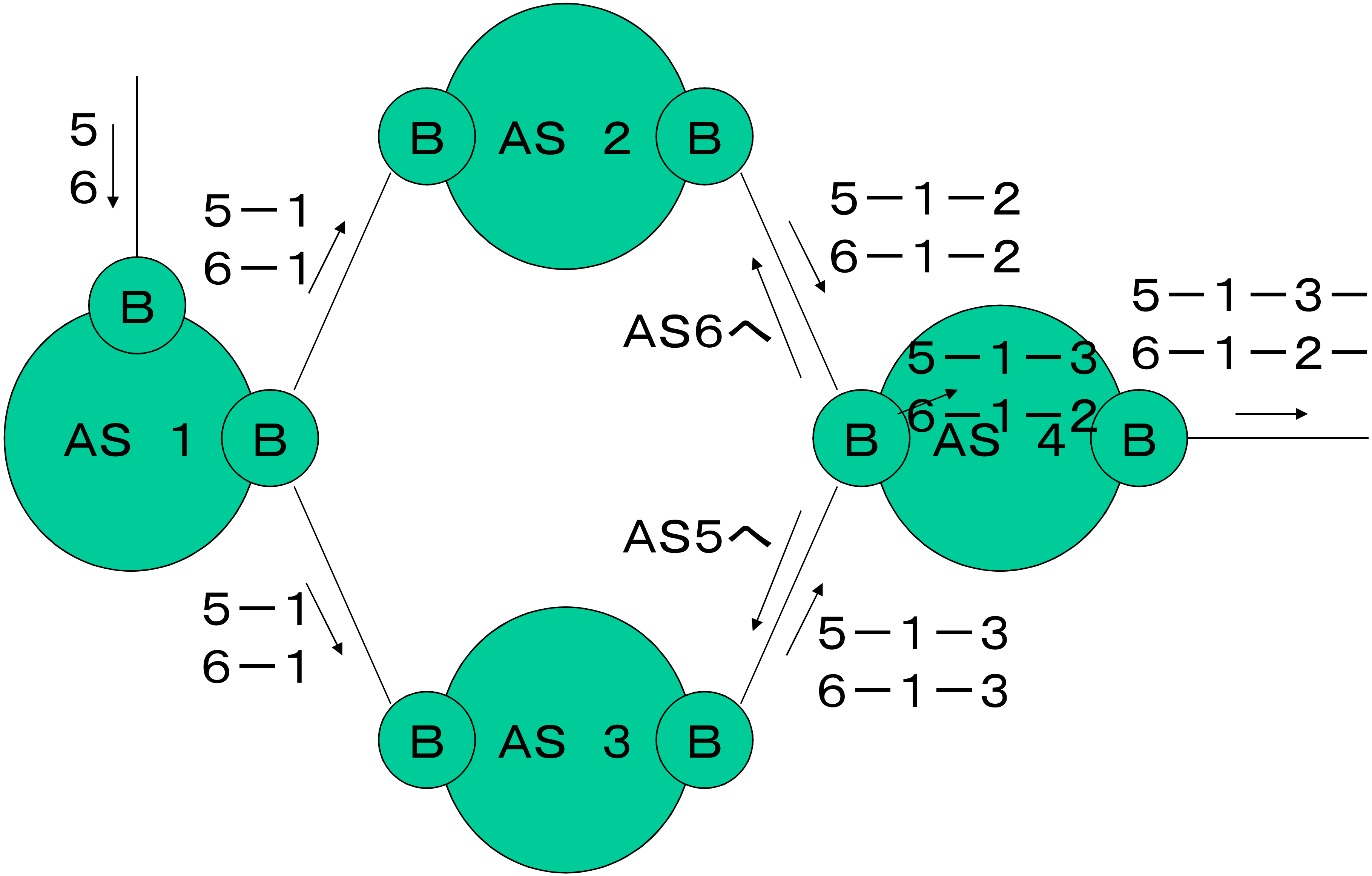
- DV style EGP
  - used between ASes (Autonomous Systems)
- instead of distance, AS path is used
  - routing entry contains address ranges belonging to AS and AS path (list of AS#s the entry passed)
  - policy determines which AS path is prefered if multiple paths exist to a destination
    - in practice, for long AS, shortest AS is often chosen
  - converge quickly, because AS path is loop free?
- border routers communicate with TCP



AS path and policy (AS4 prefer AS3 than AS2)



AS path and policy (AS2 won’t carry traffic to AS1)



AS path and load distribution

Routing Registry

- wrong address range causes world wide damage
- routing registry maintains database on which address range belongs to which AS
  - can check wrong configuration of BGP
    - wrong information won't be relayed
- a address range only belongs to a single AS
  - difficulty with anycast?

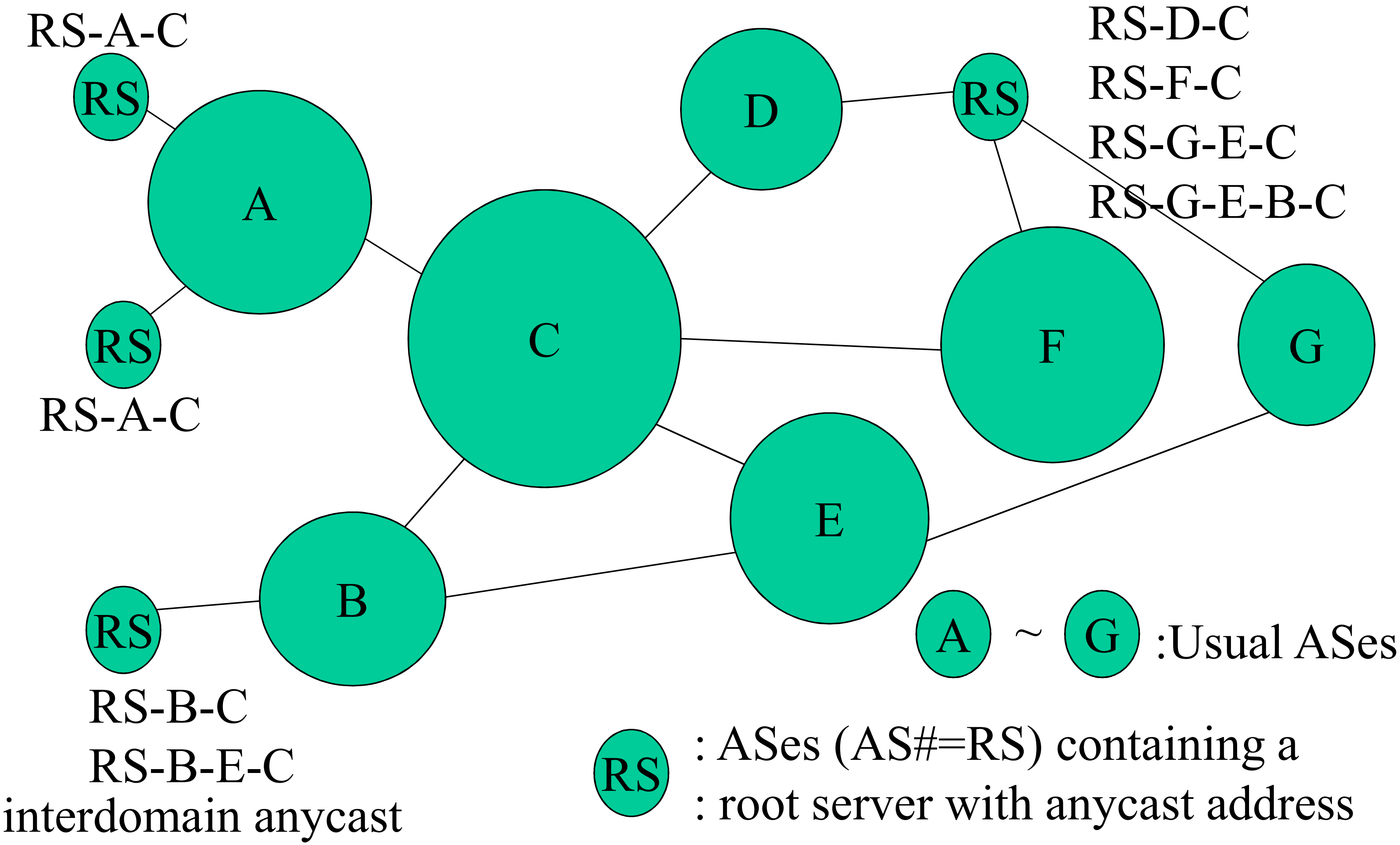
Anycast

- one address is shared by multiple servers
- packet destined to anycast address is delivered to a server determined by routing protocol
  - with IGP, nearest
  - with EGP, controllable by policy
    - use anycast AS# (shared by multiple ASes)
- a routing table entry is consumed by each anycast address

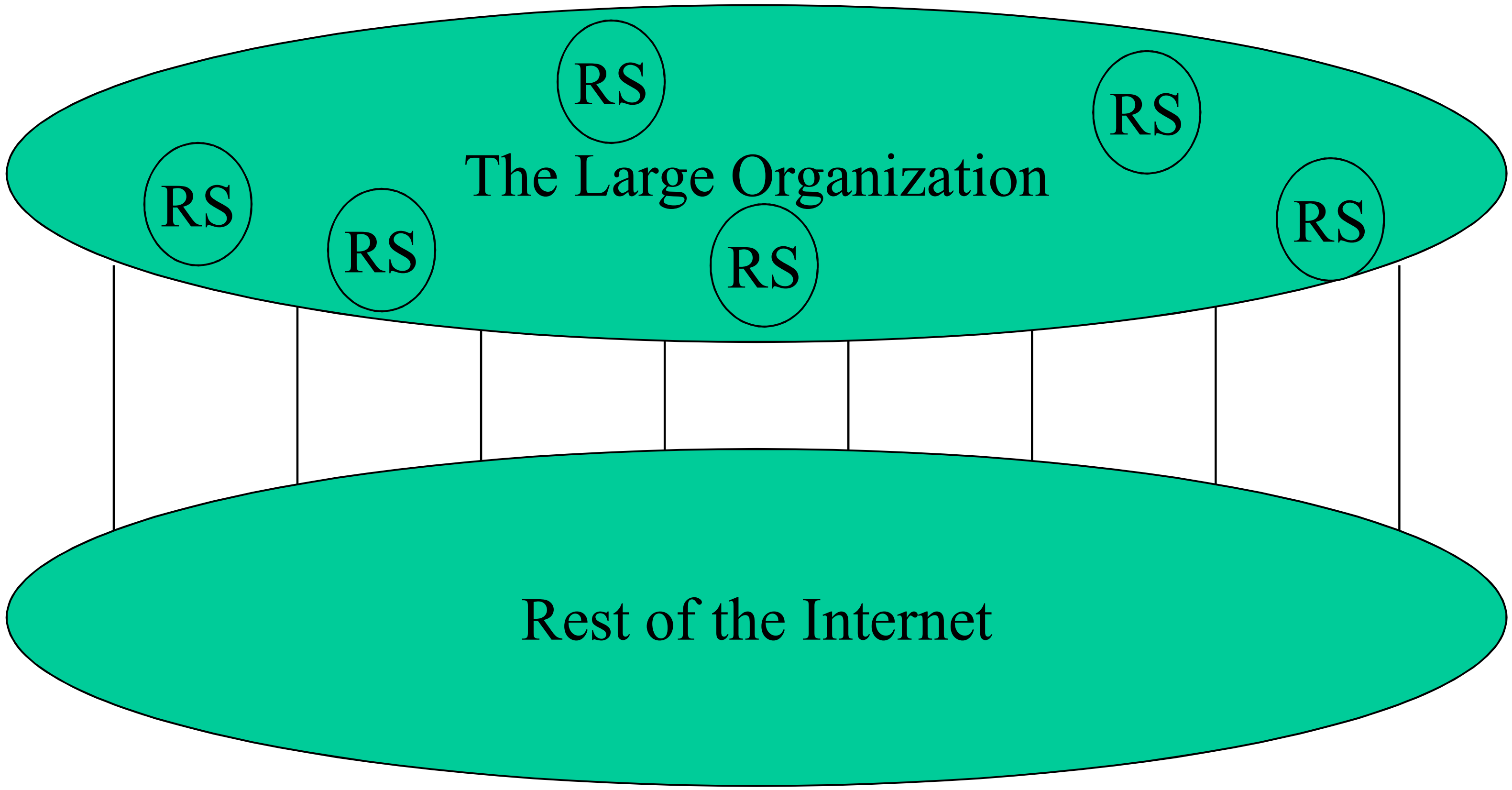
Application of Anycast

- distributed (DNS root) server
- emergency call
- get geographic location

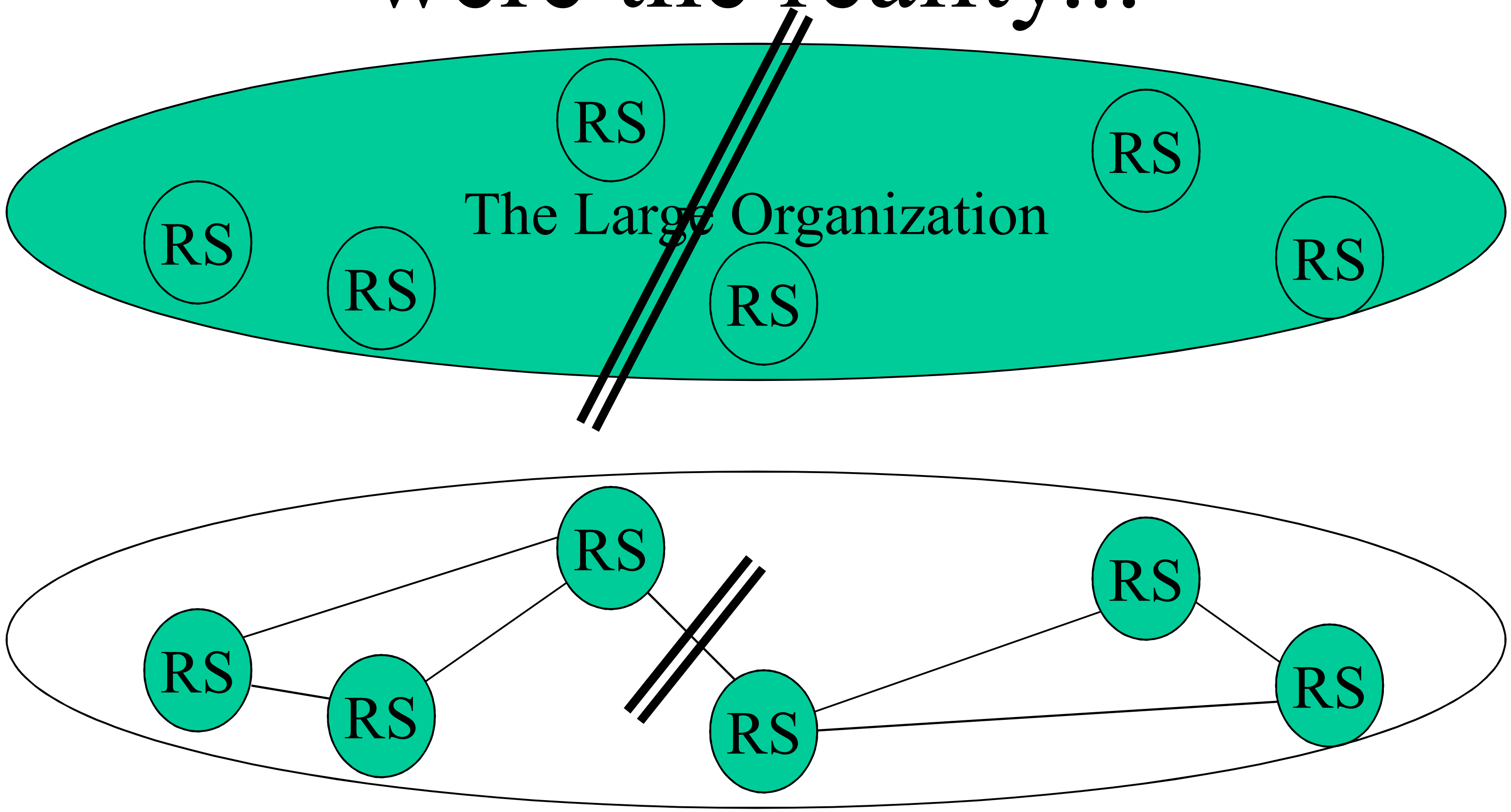
An Example (AS-pathes at C)



Another Approach: Intradomain Anycast with a Large AS

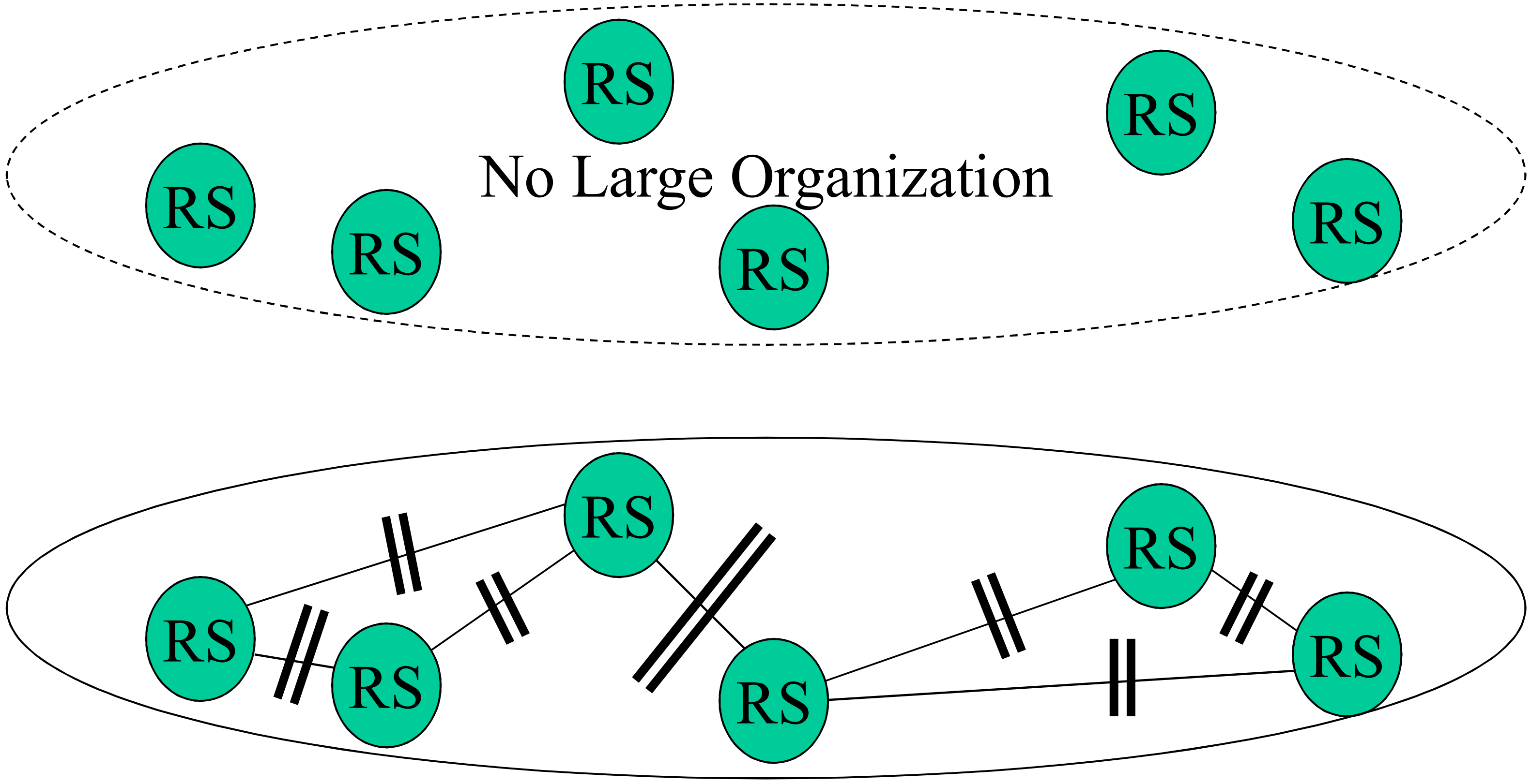


Even if the large organization(s) were the reality...

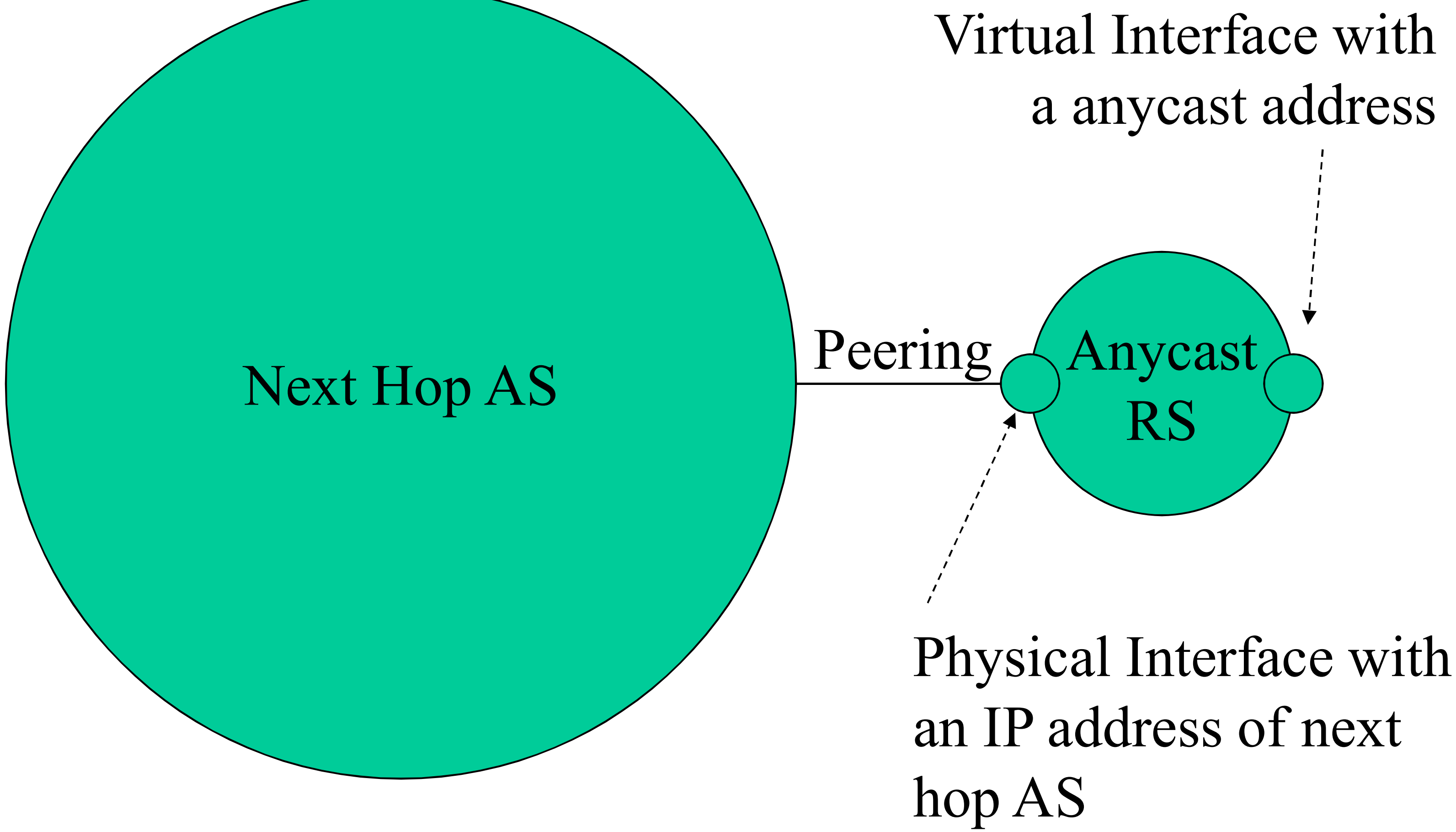


Can the geographically and topologically large organization have rich and robust internal connectivity?
Or, does the organization advertises a route to RSes only? But...

The Extreme Case



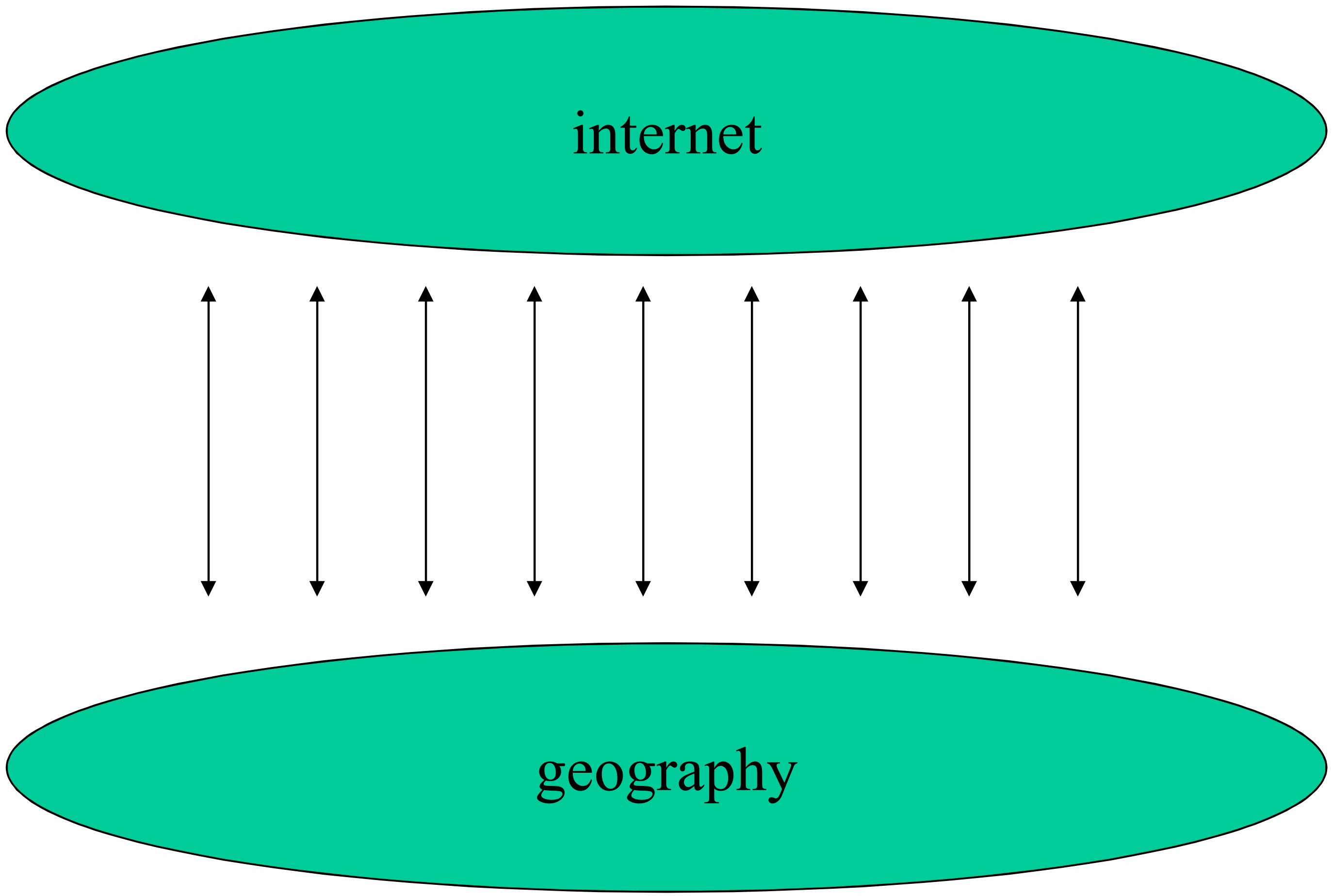
Anycast RS with Unique Addresses

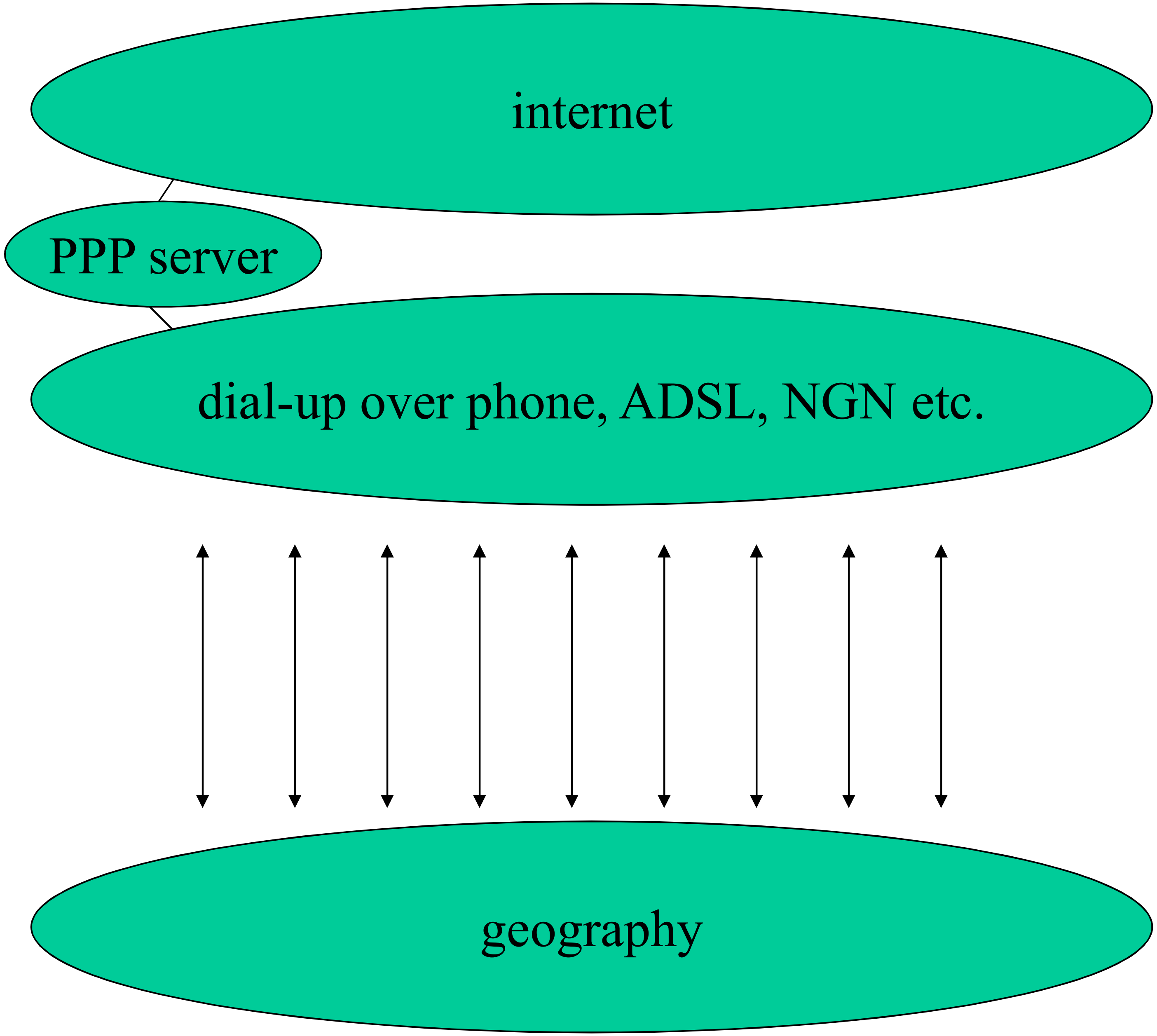


Anycast AS = Anycast Root Server

Anycast and Emergency Call

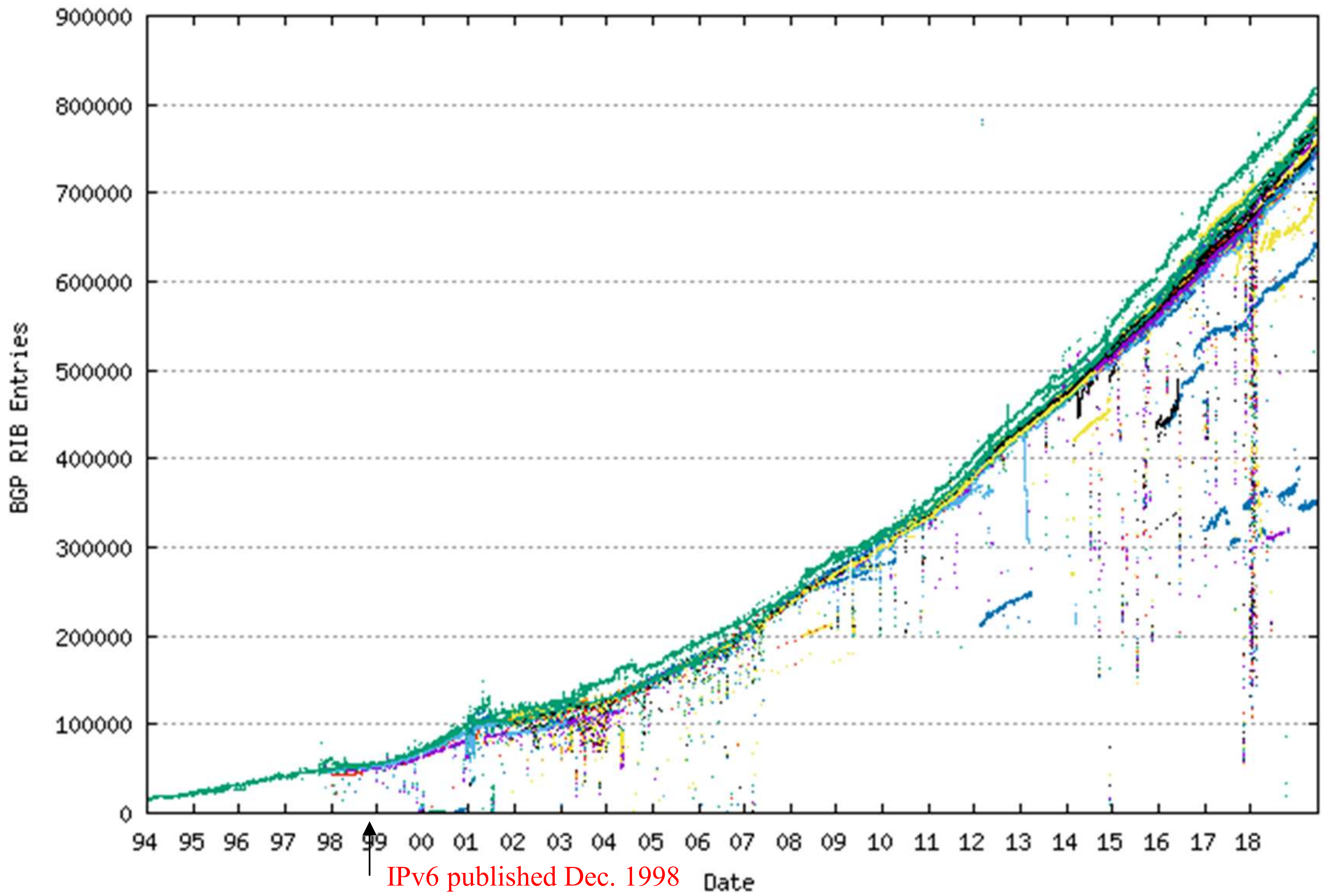
- want to communicate with nearest branch of some (police and fire) department
  - phone numbers 110 and 119 in Japanese, 911 in US
- should be possible with anycast
- with persistent internet connectivity
  - natural correspondence between topology of internet and geography
    - no correspondence with dial-up (incl. PPPoE)


internet
geography


internet
PPP server
dial-up over phone, ADSL, NGN etc.
geography

IPv4 Routing Table Size

http://bgp.potaroo.net/



Default Route

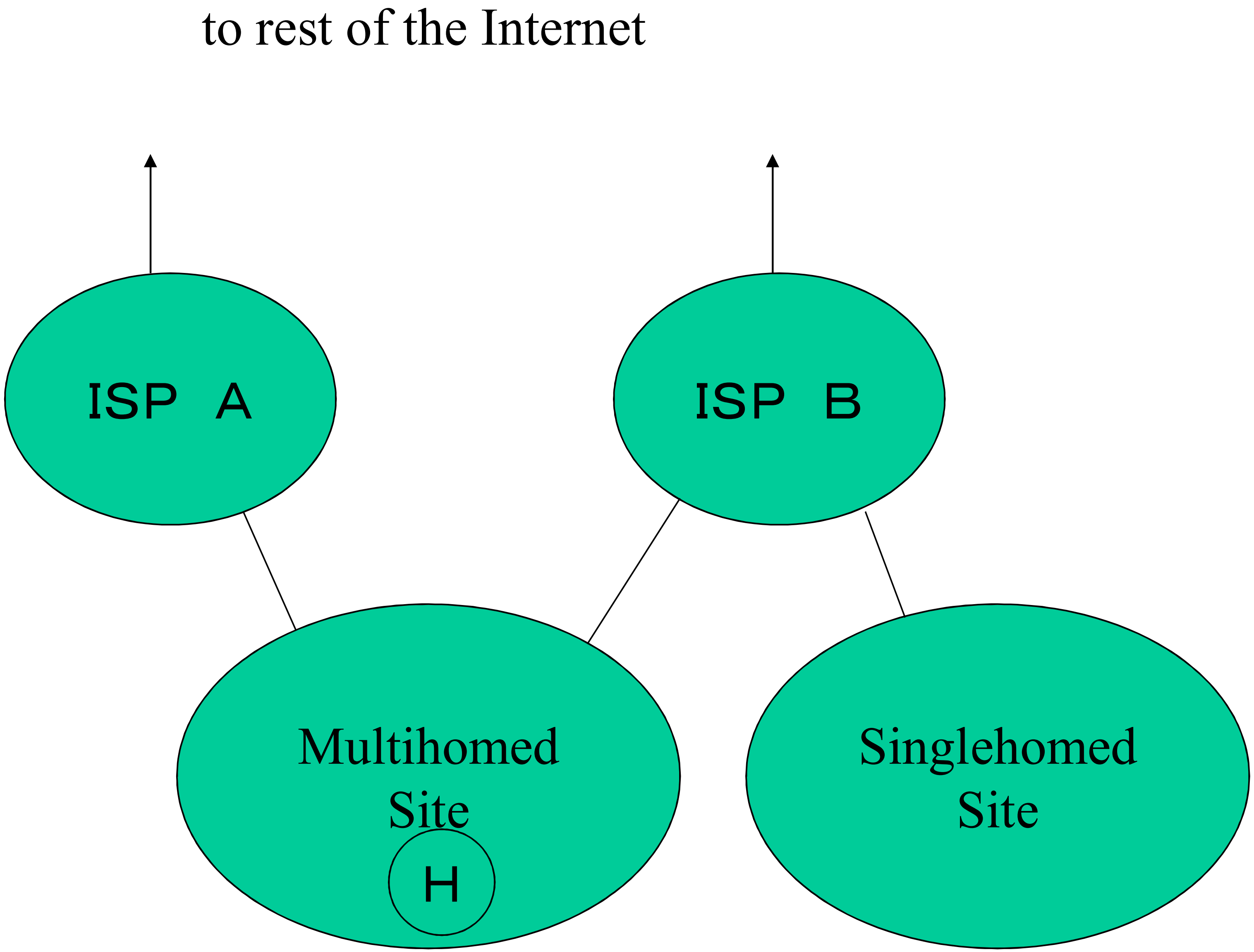
- size of global routing table
  - about 800k entries (2019)
    - steadily increasing due to multihoming by routing seemingly exponentially
- not all router must have all the route
  - can’t send entire table over slow links
- default route is used at the edge

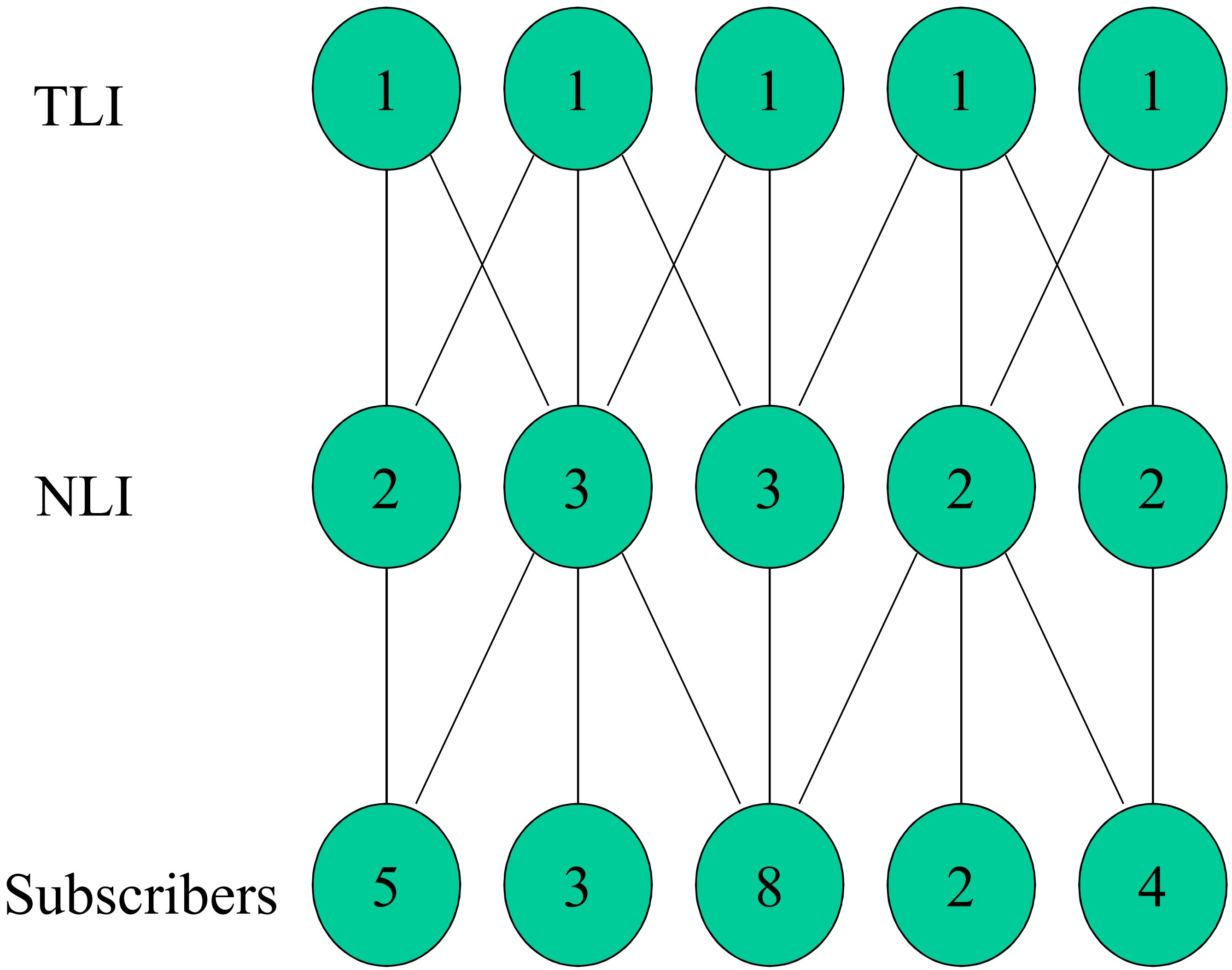
Default Router

- knowledge on IGP necessary to get route
- usual host do not have to have route?
  - separation of hosts and routers (IPv6)
  - no knowledge on IGP necessary on hosts
- all packets are sent to a nearest router
  - default router
- default router may generate ICMP redirect to refer other routers

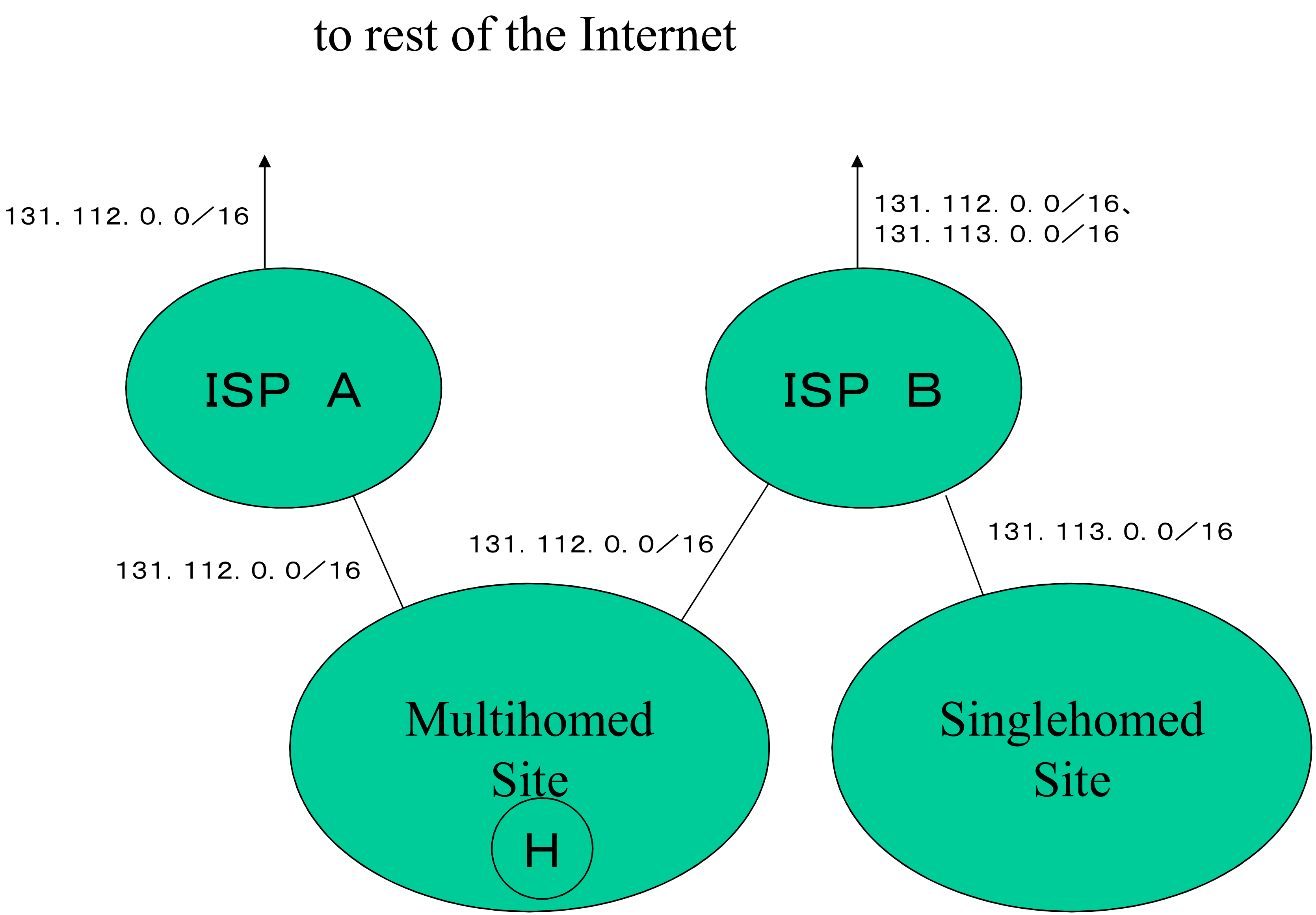
Multihoming

- have multiple upstream ISPs
  - safe even if some ISPs fail
- necessary for reliable service (incl. ISP)
  - IPv6 NLISP want to have multiple TLISPs
- multihoming by routing assumes single address with single TLA regardless of TLISP changes


to rest of the Internet
ISP A
ISP B
Multihomed
Site
H
Singlehomed
Site



Number of Prefixes with E2E Multihoming

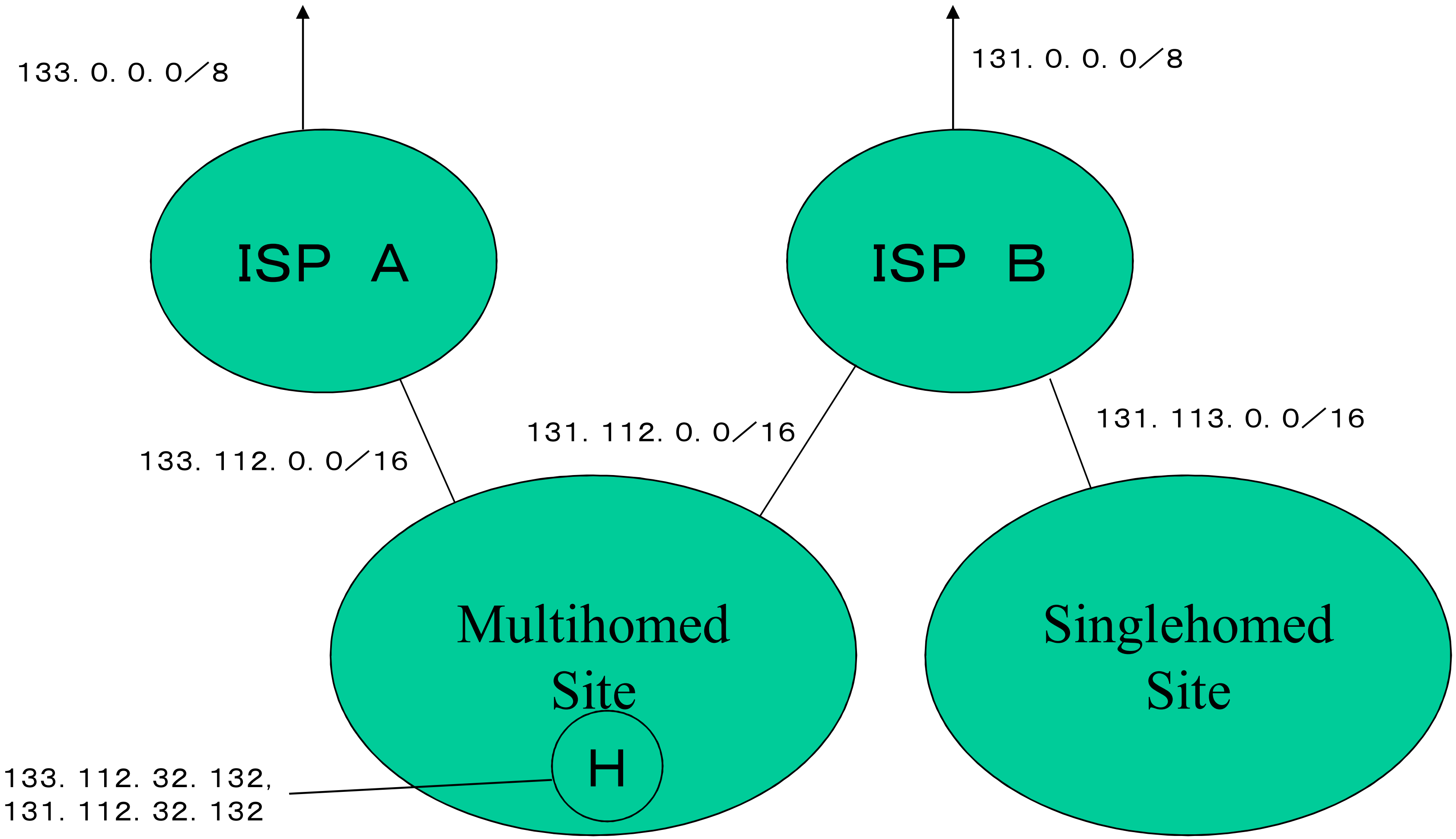


multihoming by routing

End to End Multihoming

- a host has multiple IP addresses
- peer of a host try to use multiple addresses of the host
  - rough unreachability by global routing table
  - if some address works, communication starts
  - if timeout occurs, other addresses are tried
- multihoming by routing is not necessary

to rest of the Internet



end to end multihoming

Multihoming and Default Route

- default route in multihomed AS is not very meaningful
  - exit router shouldn't be a single point of failure
- full route information is useful for E2E multihoming to choose best address of peer
- according to E2E principle, hosts should have as much information as routers
  - possible with IPv6? (failed)

Future of the Internet

- primarily by optical fiber
  - overwhelmingly high speed (>>1Tbps/core)
- wireless is still necessary
  - wireless backbone (one to many)
    - broadcast internet by satellite
      - killer application should be that of broadcast network
  - wireless access (no wiring necessary)
    - mobile internet
      - killer application should be that of phone network
        - » free conversation!

Radio Waves and the Internet

- short distance (low power)
  - install many stations (not phone network of 5G)
  - mobile internet service can be realized by IP mobility
- long distance (high power)
  - radio waves are good for one to many
  - is satellite internet fast?
    - fast only for one to many

The Mobile Internet

- mobile phone network is phone network
  - ￥0.3/128B means ￥20/sec @ 64kbps
- radio stations connected to wired high speed inexpensive flat rated internet service
  - wireless high speed inexpensive flat rated internet
  - security improvement necessary (802.11ai)
- wireless internet + IP mobility = the mobile internet

The Mobile Internet

- wireless Internet + IP mobility
  - free movement around a single station by wireless communication
  - IP mobility keeps same IP address and TCP connection upon station changes

Wireless Internet

- needs wired Internet infrastructure
  - by densely installed optical fiber
    - FCC once claimed wireless only is enough, but,
  - high speed inexpensive radio stations attached to wired high speed inexpensive flat rated internet
    - inexpensive flat rated wireless internet
  - if stations are dense enough
    - high speed inexpensive flat rated wireless internet

Technical Problems of the Wireless Internet

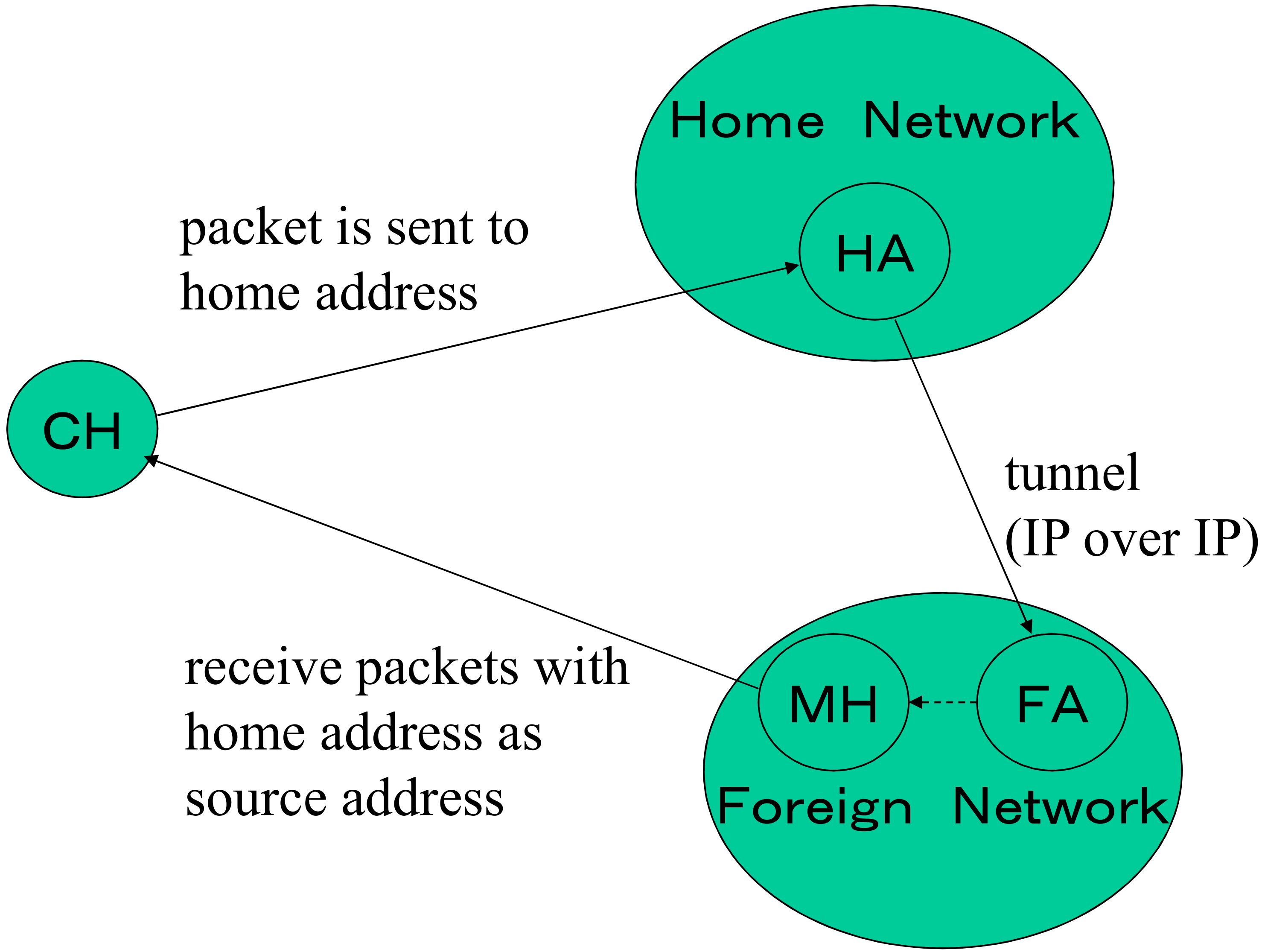
- wireless can be used by general public
  - authentication
    - good that anyone can use the internet anytime/anywhere
    - no good if users are not identified
      - crime investigation
      - charge money
  - encryption
    - basically should be end to end
    - good for old protocols with plain text password

Security of Wireless Internet

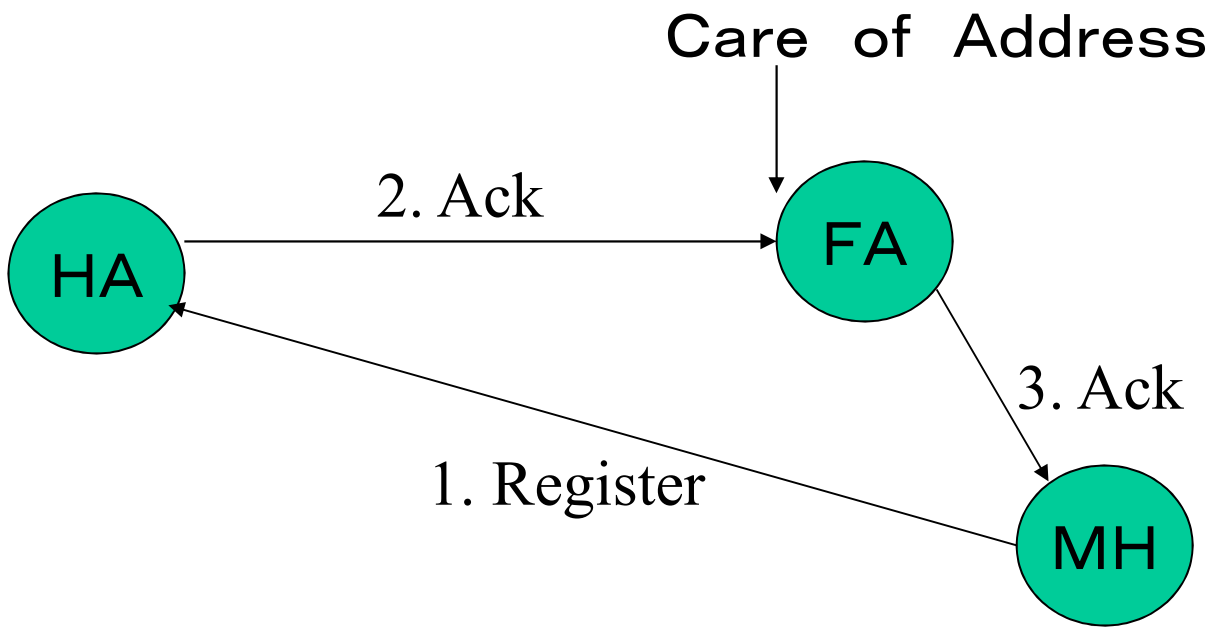
- user key is managed by RADIUS server
- user generate session key, encrypted the key by user key and send the key to base station
- base station ask RADIUS server decrypt the session key
- session key may be used for packet-wise authentication and encryption

IP Mobility (rfc2002)

- want to keep using same IP address even if location of hosts change
  - keep TCP connections
- four elements
  - HA (Home Agent)
  - FA (Foreigh Agent)
  - MH (Mobile Host)
  - CH (Correspondent Host), usual host



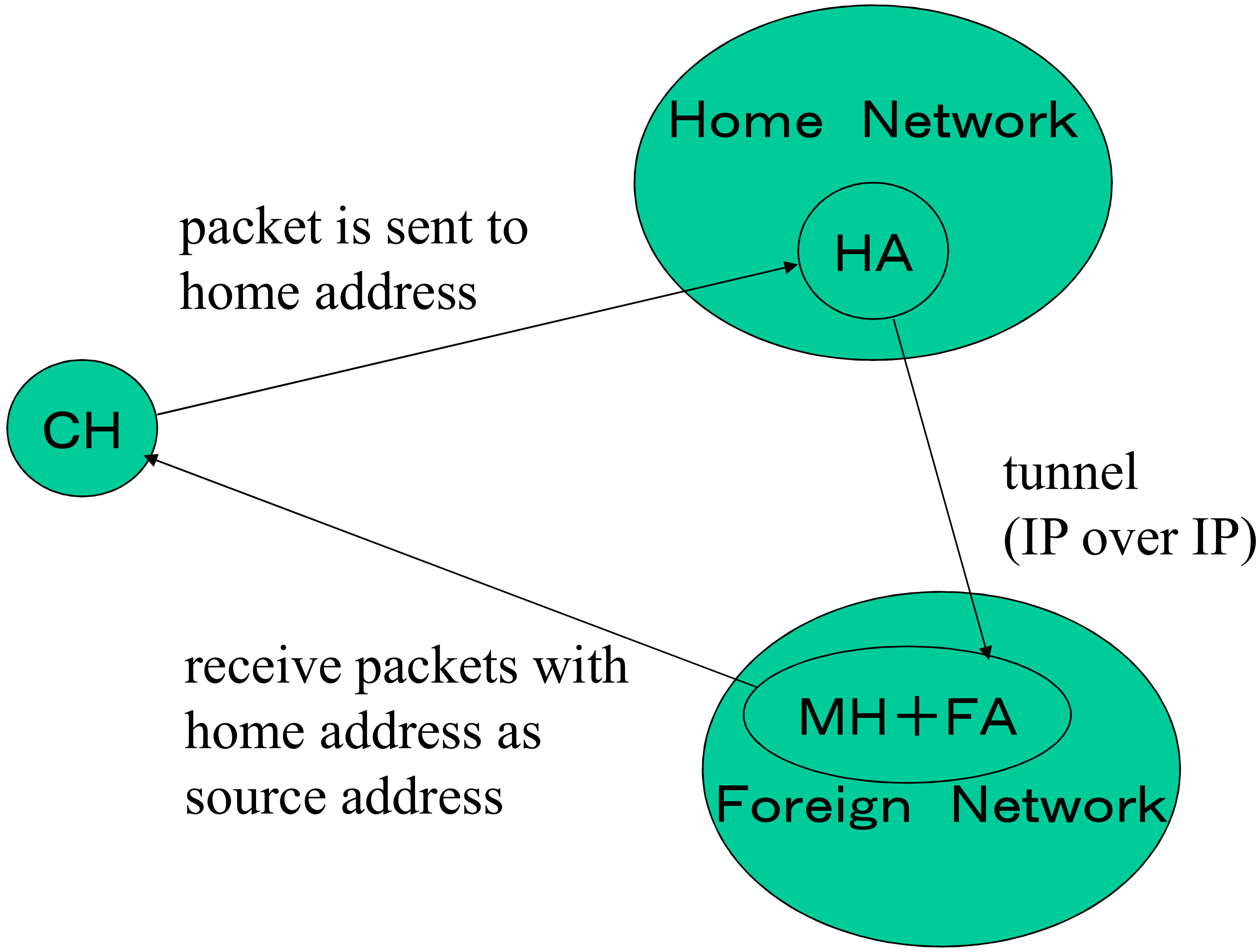
triangular exchanges of packets



Registration of Care of Address to HA

Ends of Mobility?

- HA
  - maintain location of MH (as foreign address)
  - forward packets to FA
- MH
  - register foreign address to HA
- FA?
  - intermediate device between HA and MH
  - against E2E principle?



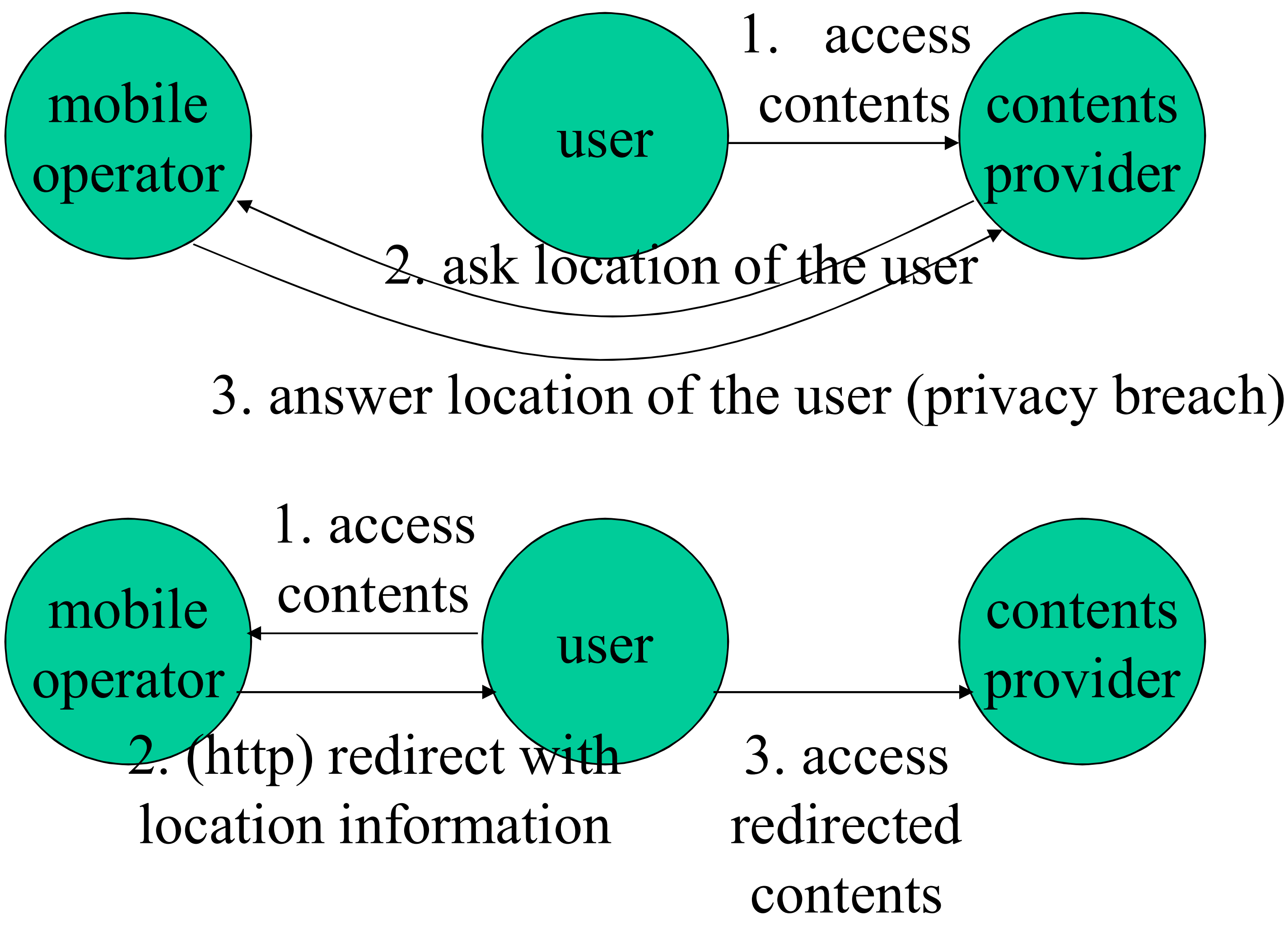
inclusion of FA by MH

Offering Location Dependent Contents by Anycast

- with mobile internet
  – all the usual internet contents may be offered
  – may also offer location dependent contents
- by sharing anycast address by base stations
  – base stations acting as contents server
    - different contents by base station
- if contents is accessed by anycast address
  – contents of nearest base station is offered

Location Dependent Contents and Privacy

- mobile operator knows location of users
  - mobile operator can offer location dependent contents
- if mobile operator tells location of users
  - to other contents providers, it is privacy breach
    - other contents providers cannot offer location dependent contents
  - to users is not a problem
    - the users may, further, tell content providers their locations



mobile operator
user
1. access contents
contents provider
2. ask location of the user
3. answer location of the user (privacy breach)
1. access contents
mobile operator
user
contents provider
2. (http) redirect with location information
3. access redirected contents

Wrap-up (1)

- routing protocol is intelligence of network
  - with DV, network compute route distributed way
  - with LS, network merely distribute information
  - routers (ends of routing protocol) perform more function with LS than DV
    - BGP perform distributed computation, but, the computation must be based on local policies of ASes
- anycast can be useful

Wrap-up (2)

- default route makes routers more intelligent than hosts
  - against E2E principle
- multihoming by routing compute route to each site by network
  - not scale, end to end multihoming necessary
- mobility needs security (authentication)
- FA of MIP is intelligent intermediate entity
  - against E2E principle