

インターネットインフラ特論

4. IPv4、ICMP、ARP

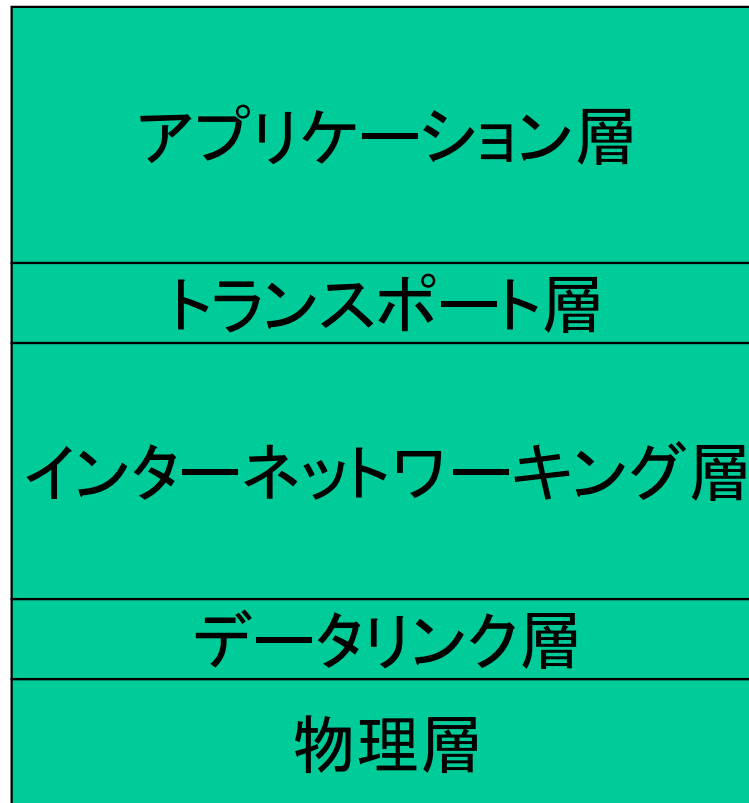
太田昌孝

mohita@necom830.hpcl.titech.ac.jp

<ftp://chacha.hpcl.titech.ac.jp/infra4j.ppt>

インターネットのレイヤリング

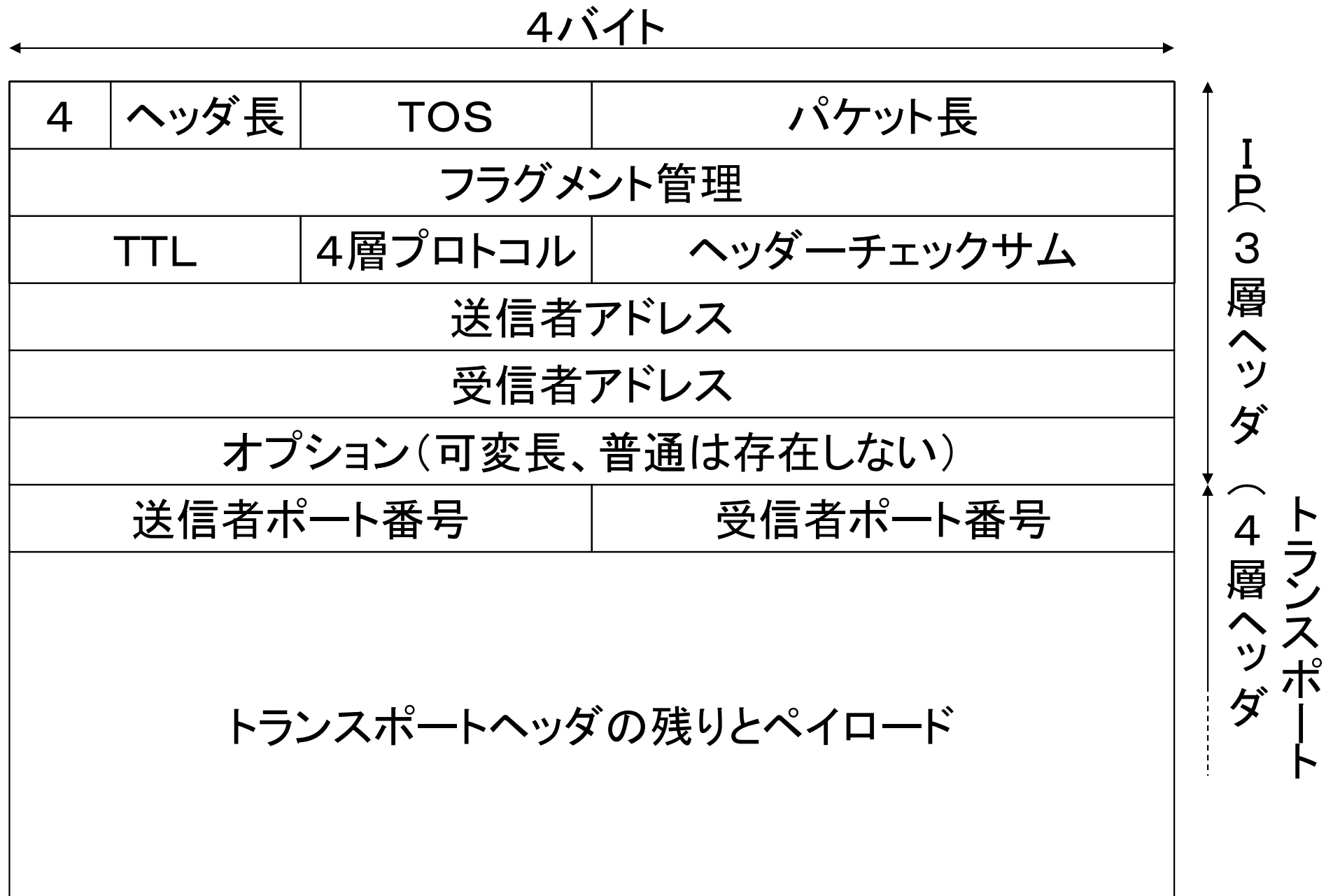
- 物理層、アプリケーション層は必須
- インターネットワーキング層はできる限りのことをやる
- データリンク層、トランスポート層は極力なにもやらない



インターネットのレイヤリング構造

IPv4 (Internet Protocol Version 4、RFC791)

- ネットワーク中ではほとんど何もやらない
 - パケットの目的地への配送
 - フラグメンテーション
 - TTLの管理



IPv4パケットフォーマット(RFC791)

ヘッダ長

- 4ビットフィールド
- 4バイト単位でのヘッダの長さ
 - 最少5(20バイト)、最大15(60バイト)、実質5のみ

TOS (Type Of Service)

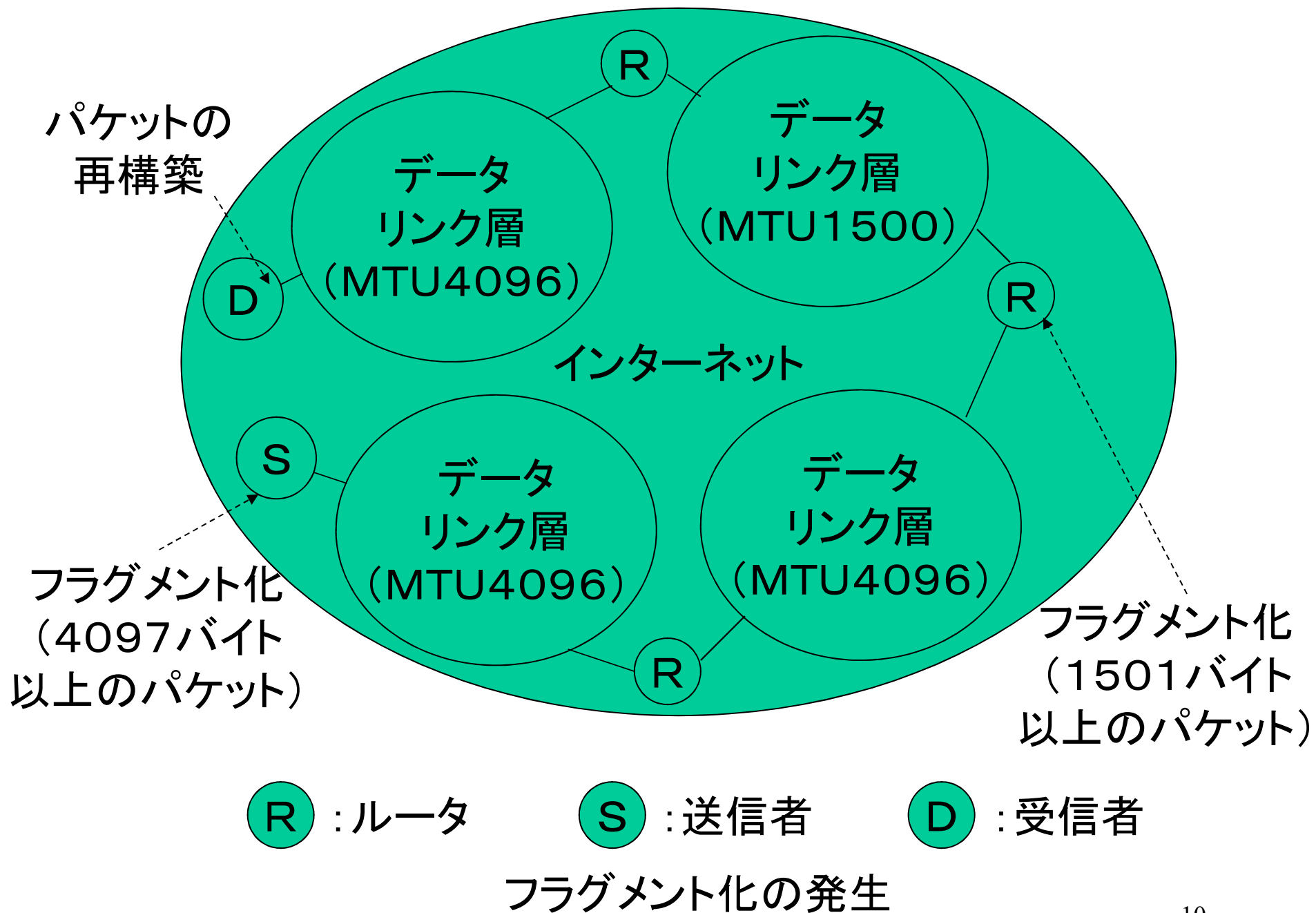
- 8ビットフィールド
- パケット優先度
- 使われていない
- DiffServeやECNとして再利用の動き

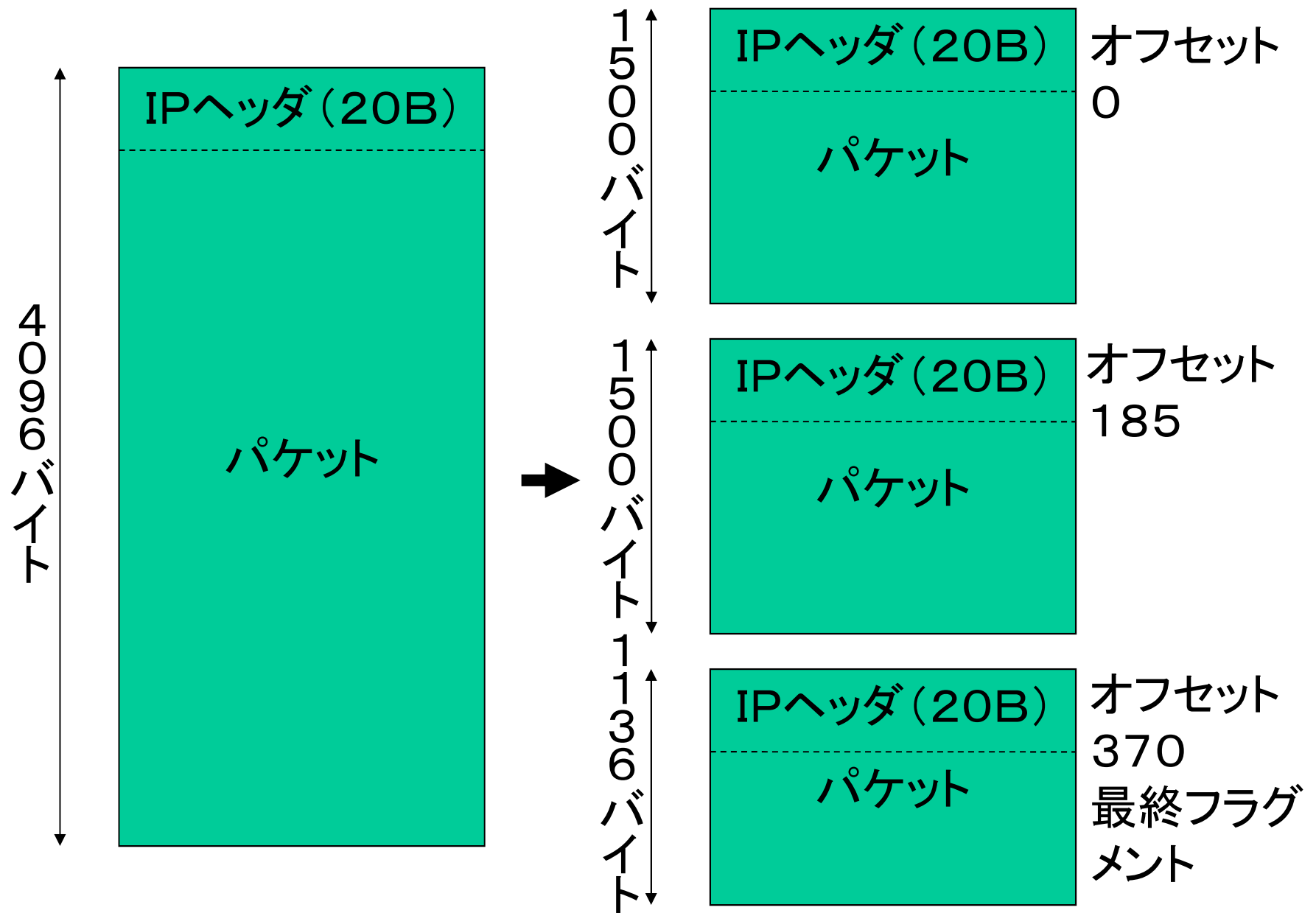
パケット長

- 16ビットフィールド
- バイト単位のパケットの長さ
- 最少20、最大65535
- どんなホストでも受け取れるのは576バイト(ヘッダを除くと516バイト)まで

フラグメント管理

- MTUより大きいパケットは分割して送る
 - 最少MTUは68バイト
- 16ビットIDフィールド
 - フラグメント
- 3ビットフラグフィールド
 - フラグメント不許可、フラグメント化など
- 13ビットオフセットフィールド
 - 8バイト単位でのパケット中のフラグメント位置

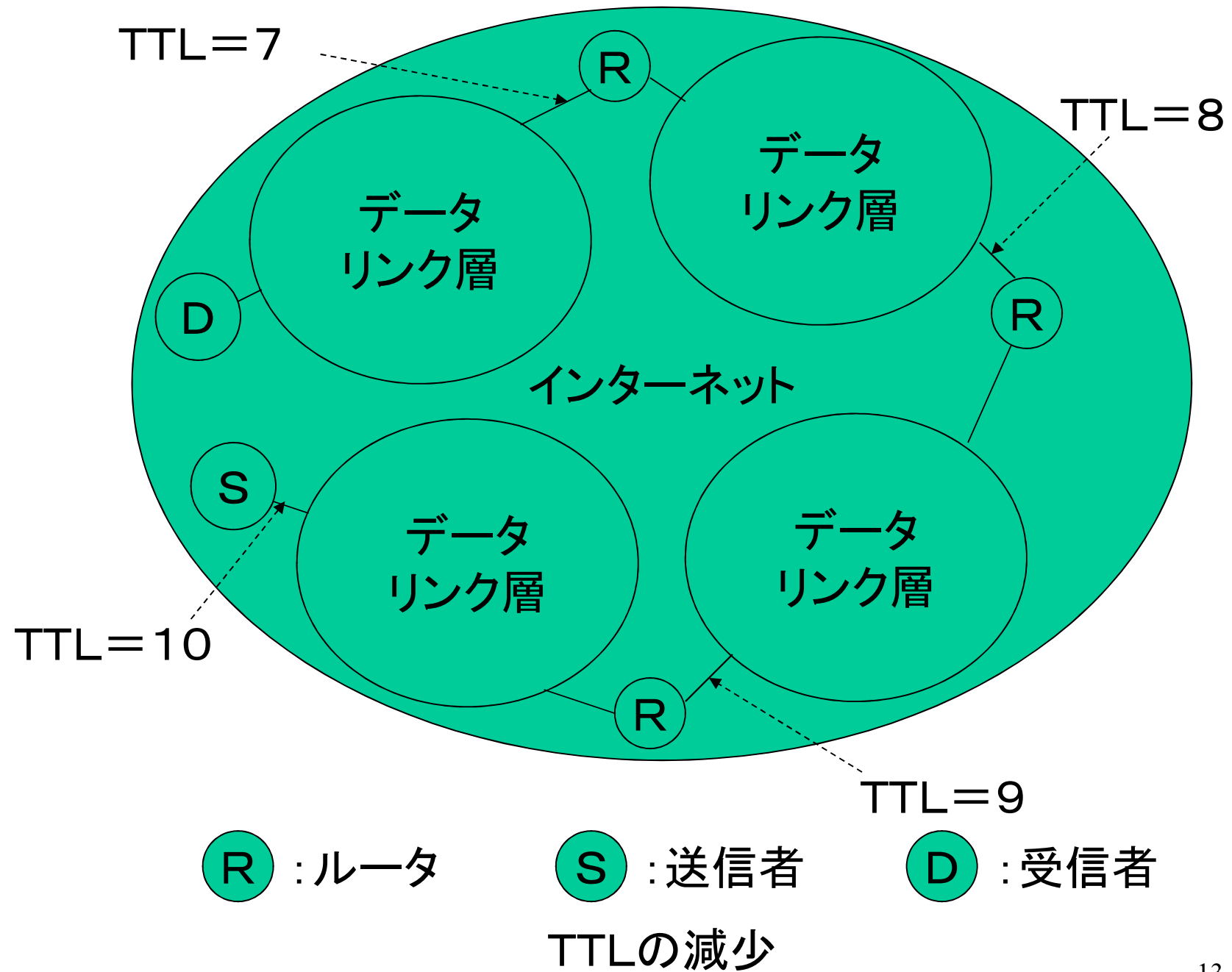




パケットフラグメント化のあまり良くない例 (MTU1500)¹¹

TTL (Time To Live)

- 8ビットフィールド
- パケットの寿命を管理
 - ルータを経由するたびに少なくとも1減る
 - ルータ中にN秒滞在すればN減る
 - TTLが0になったパケットは破棄
 - パケットの無限ループを防止



4層プロトコル

- 8ビットフィールド
- トランスポートプロトコルを区別
 - ICMPは1
 - TCPは6
 - UDPは17

ヘッダーチェックサム

- 16ビットフィールド
- ヘッダーを16ビット単位で加算(1の補数)したもの

送信者アドレス

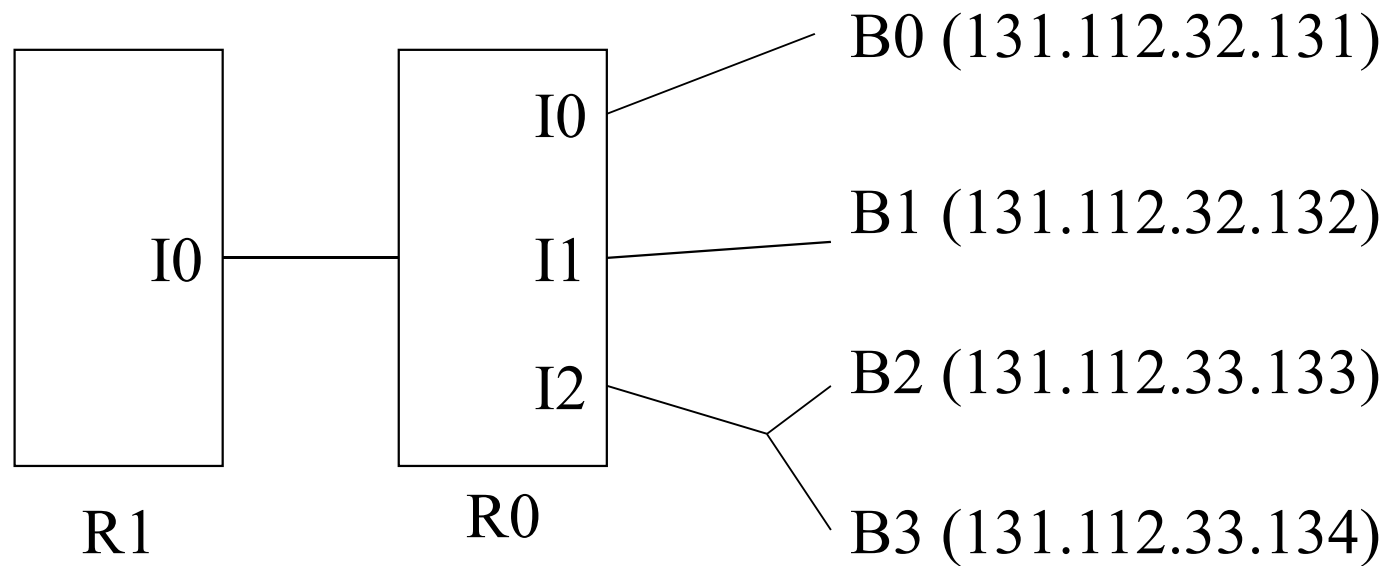
- パケット送信者のアドレス
- パケット返信に利用
 - そのままでは信用できない
- ポート番号と組み合わせて、トランスポート層での相手の識別にも利用

受信者アドレス

- パケット受信者のアドレス
- ルータは、受信者アドレスをキーに経路表を検索してパケットの経路を決定
 - 経路表の検索には、受信者アドレスに可変長のマスクをつける
- ホストは(ルータも)、自分の受信者アドレスの場合、そのパケットを受け取る

経路表 (Routing Table)

- ルータは経路表の検索結果に基づきパケットを次のルータに送出
 - 経路表は受信者アドレスで引く
- 似たIPアドレスを持つホストがある地域にだけあれば、経路表エントリは1つで済む
 - 経路の縮約 (Route Aggregation)
 - 経路表はアドレスパターンごとに1エントリ
 - 電話でも、国、市外局番、市内局番と階層化
 - +81-3-5734-3299



R0の経路表

目的地	次
131.112.32.131	I0
131.112.32.132	I1
131.112.33.*	I2

R1の経路表

目的地	次
131.112.*	I0

経路表の縮約

クラス別ルーティング

- IPv4アドレスを5つのクラスに分類
 - クラスA、B、Cはユニキャストに
 - クラスDはマルチキャスト、Eはリザーブ
- ユニキャストIPアドレスを前半(ネットワーク部)、後半(ホスト部)にわけ、ネットワーク(データリンク)単位でルーティング
 - ホスト部全部1はネットワーク内ブロードキャストのアドレス
 - ホスト部全部0はネットワーク自体のアドレス

クラス別ルーティングの問題点

- 個別データリンク内のホスト数は多くても数十程度
 - クラスCでも大きすぎる
 - ルート情報の不必要な増大
 - IPv4アドレスの不必要な消費
- IPv4アドレス構造の細分が必用
 - サブネット

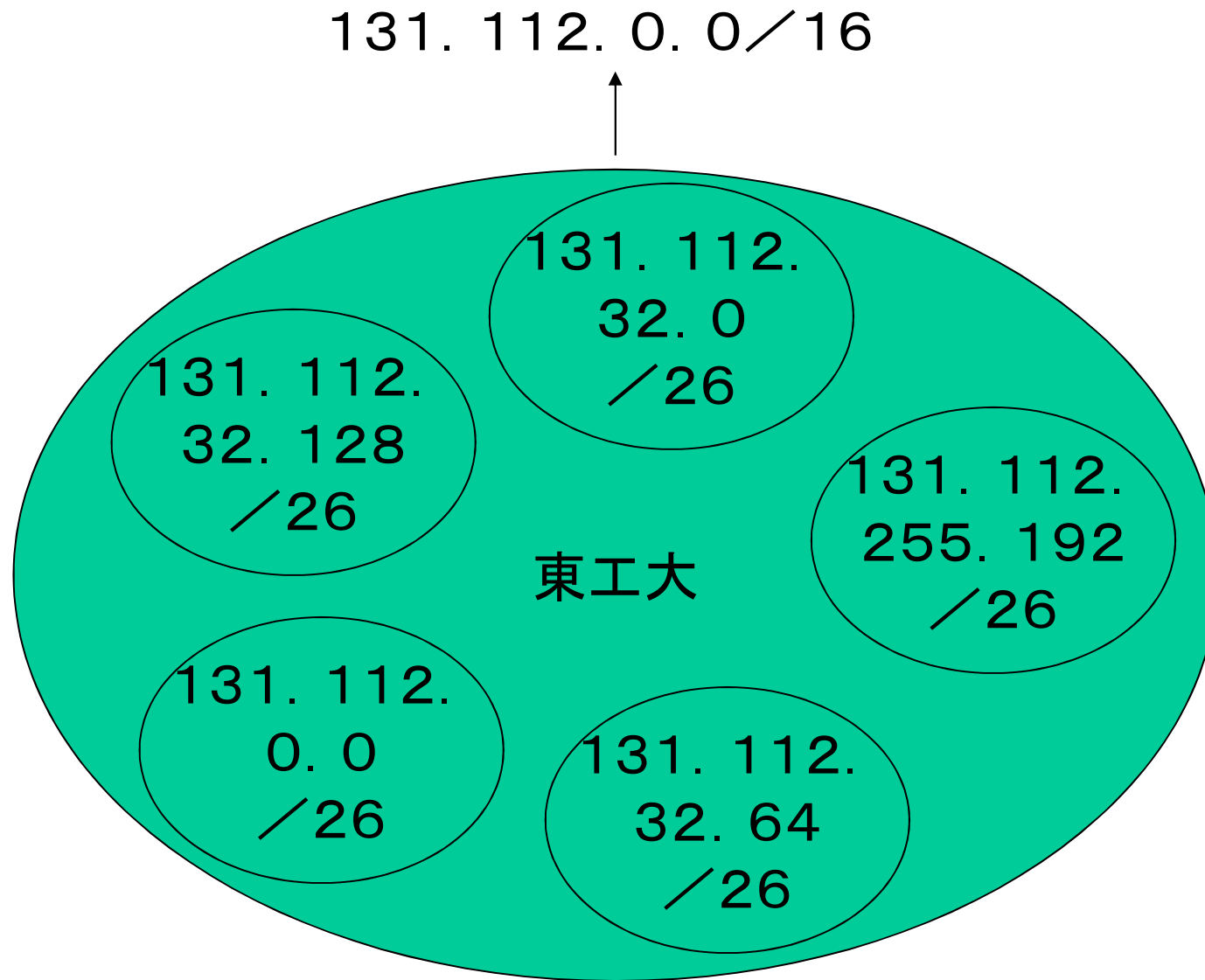
サブネット

- ユニキャストIPアドレスのホスト部を前半（サブネット部）、後半（ホスト部）にわけ
- ネットワーク内ではサブネット（データリンク）単位でルーティング
 - 1組織にクラスBアドレス1つでほぼ十分
- ネットワーク外ではネットワーク単位でルーティング
 - 外部にはルート情報は1つしかみえない



131. 112. 32. 128 / 26

サブネット化IPアドレスの構造の例(東工大)



CIDR (Classless InterDomain Routing) (RFC1519)

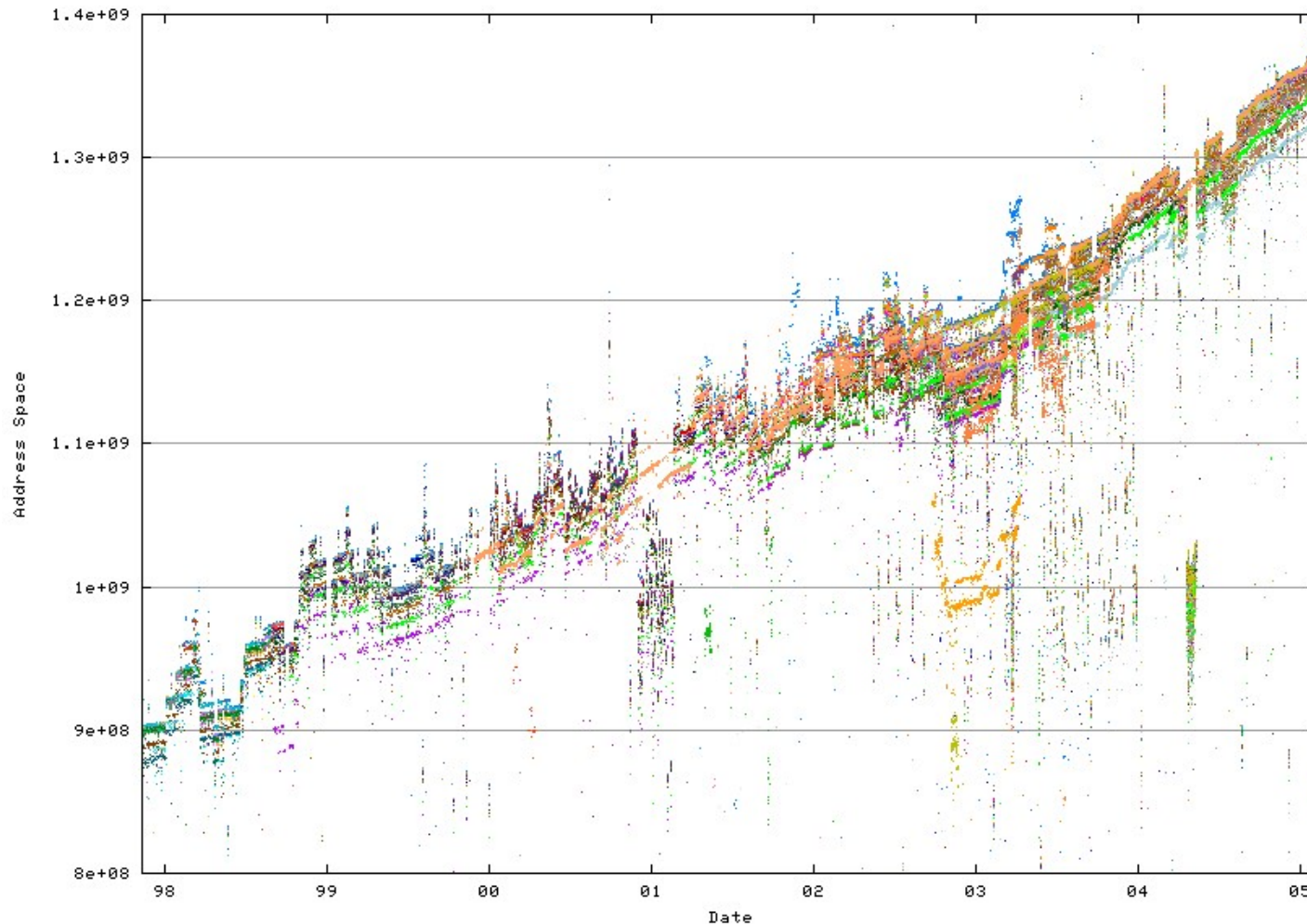
- クラスA、B、Cの区別を撤廃
 - ルーティングプロトコルはネットマスク長も運ぶ
- 階層的アドレス割り当ての例
 - ISPにアドレスを256個割り当て
 - ISP外部では／24でルーティング
 - ISPは各顧客にアドレスを8個ずつ割り当て
 - ISP内部では32個の／29を個別にルーティング

IPv4の問題点

- アドレス空間が足りない
 - 32ビットでは40億端末が限界
- 経路表が大きくなる
 - CIDR化の努力にもかかわらず、現在は37万エントリ程度(日々増加の一途)

IPv4 Address Span

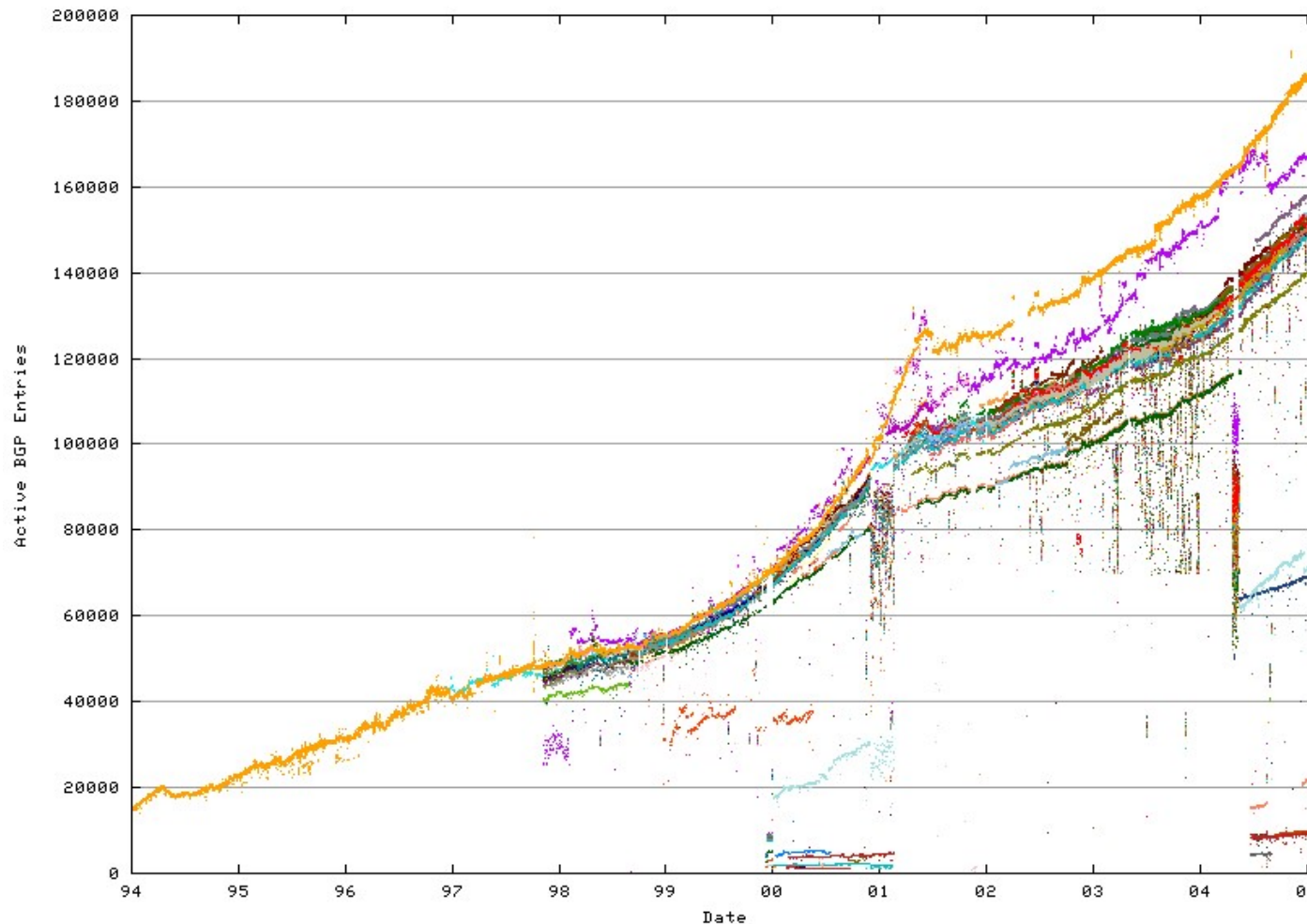
<http://www.apnic.net/meetings/19/docs/sigs/routing/routing-pres-info-huston-routing-table.ppt>



This figure shows the total amount of address space spanned by the routing table. This is a view derived from the Route-Views archive, where each AS has a single colour. The snapshots are at two-hourly intervals, and span from early 2000 until the present. The strong banding in the figure is spaced 16.7M units apart, or the size of a /8 advertisement. There appear to be 3 /8 advertisements that are dynamic. Not every AS sees the same address range, and this is long term systemic, rather than temporary. This is probably due to routing policy interaction, coupled with some cases of prefix length filtering of routing information. The rate of growth declined sharply across 2002 and the first half of 2003, resuming its 2000 growth levels in 2004.

IPv4 Routing Table Size

<http://www.apnic.net/meetings/19/docs/sigs/routing/routing-pres-info-huston-routing-table.ppt>



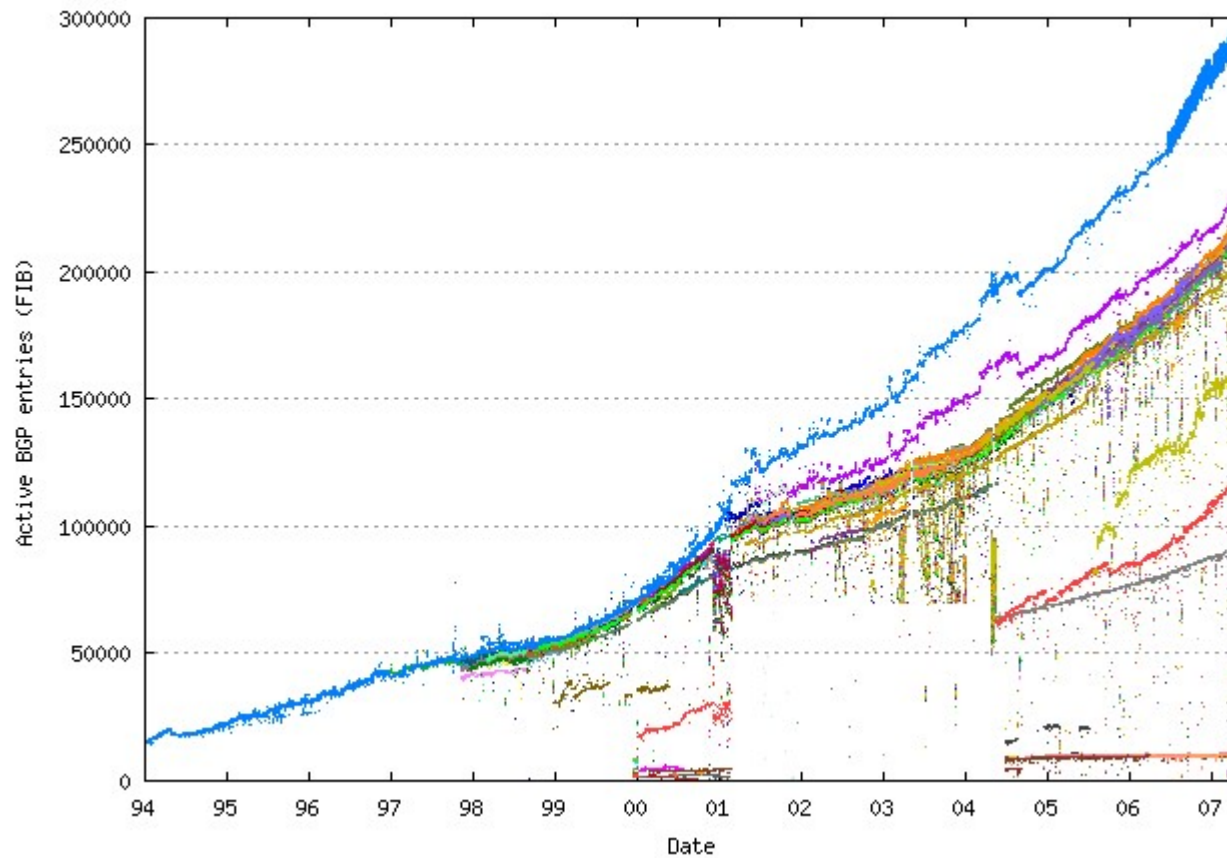
Data assembled from a variety of sources, including Surfnets, Telstra, KPN and Route Views. Each colour represents a time series for a single AS.

The major point here is that there is no single view of routing. Each AS view is based on local conditions, which include some local information and also local filtering policies about external views.

IPv4 Routing Table Size

<http://bgp.potaroo.net/>

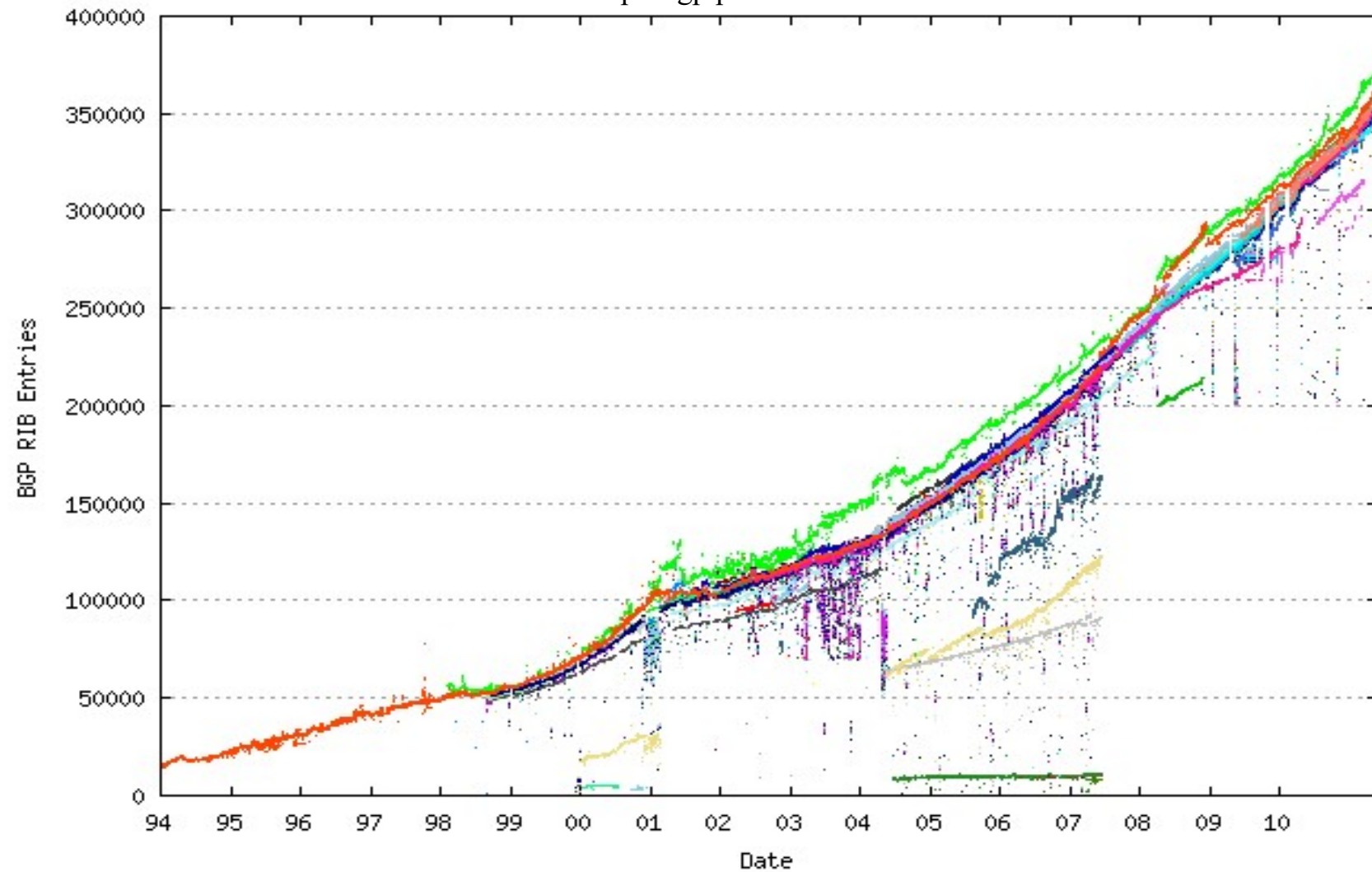
Growth of the BGP Table - 1994 to Present



(Data Gathered from AS1221 and Route-Views)

IPv4 Routing Table Size

<http://bgp.potaroo.net/>



経路の縮約が不可能な場合

- 経路がアドレスパターンで決まれば縮約可
- 経路が受信者アドレスだけで決まらない
 - QoSルーティング
- 受信者アドレスが場所と関係ない
 - マルチキャスト
- 地域のIPアドレスにまとまりがない
 - IPv4
 - ルーティングによるマルチホーミング

ICMP (Internet Control Message Protocol、RFC792)

- 制御やエラーのためのプロトコル
- IP層のためのものとトランスポート層のためのものが混在
- パケットエラーに対するものは
 - 64ビットのICMPヘッダ + ICMPの原因となったパケットのIPヘッダとそれ以降の64ビット

ICMPメッセージの種類

- Destination Unreachable
- Time Exceeded
- Parameter Problem
- Source Quench
- Redirect
- Echo & Echo Reply
- Time Stamp & Time Stamp Reply
- Information Request & Reply

ICMP

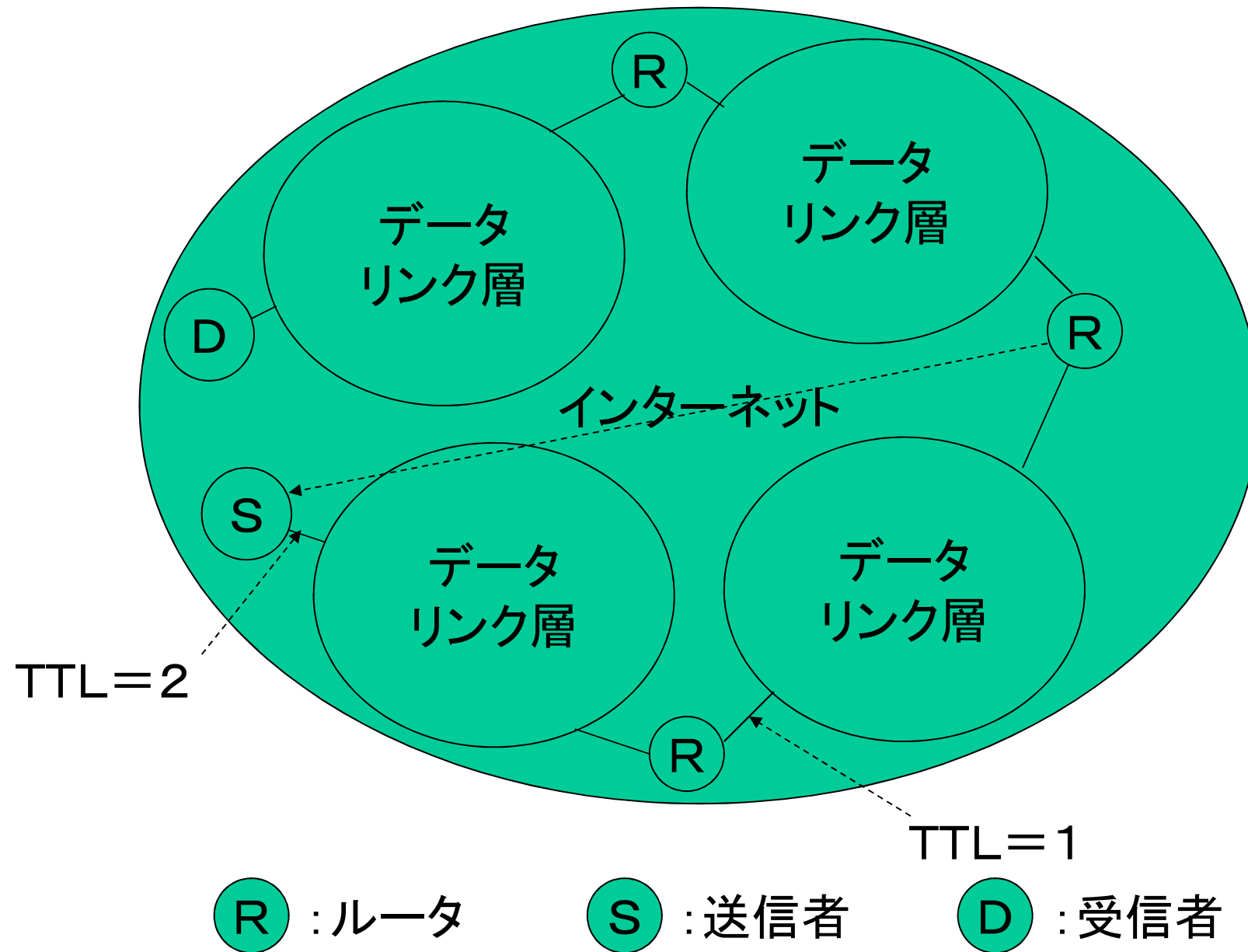
Destination Unreachable

- 各種の到達不能性
 - ネットワークへの経路がない
 - ホストへの経路がない
 - 相手ホスト上でプロトコルがサポートされない
 - 相手ホスト上でポートがサポートされない
 - フラグメントが必要だが不許可
 - ソースルートが失敗

ICMP

Time Exceeded

- TTLが0になった
 - TTLを少しずつ増やし、ICMP Time Exceededがどこからくるか見れば、パケットの経路を調べられる(traceroute)



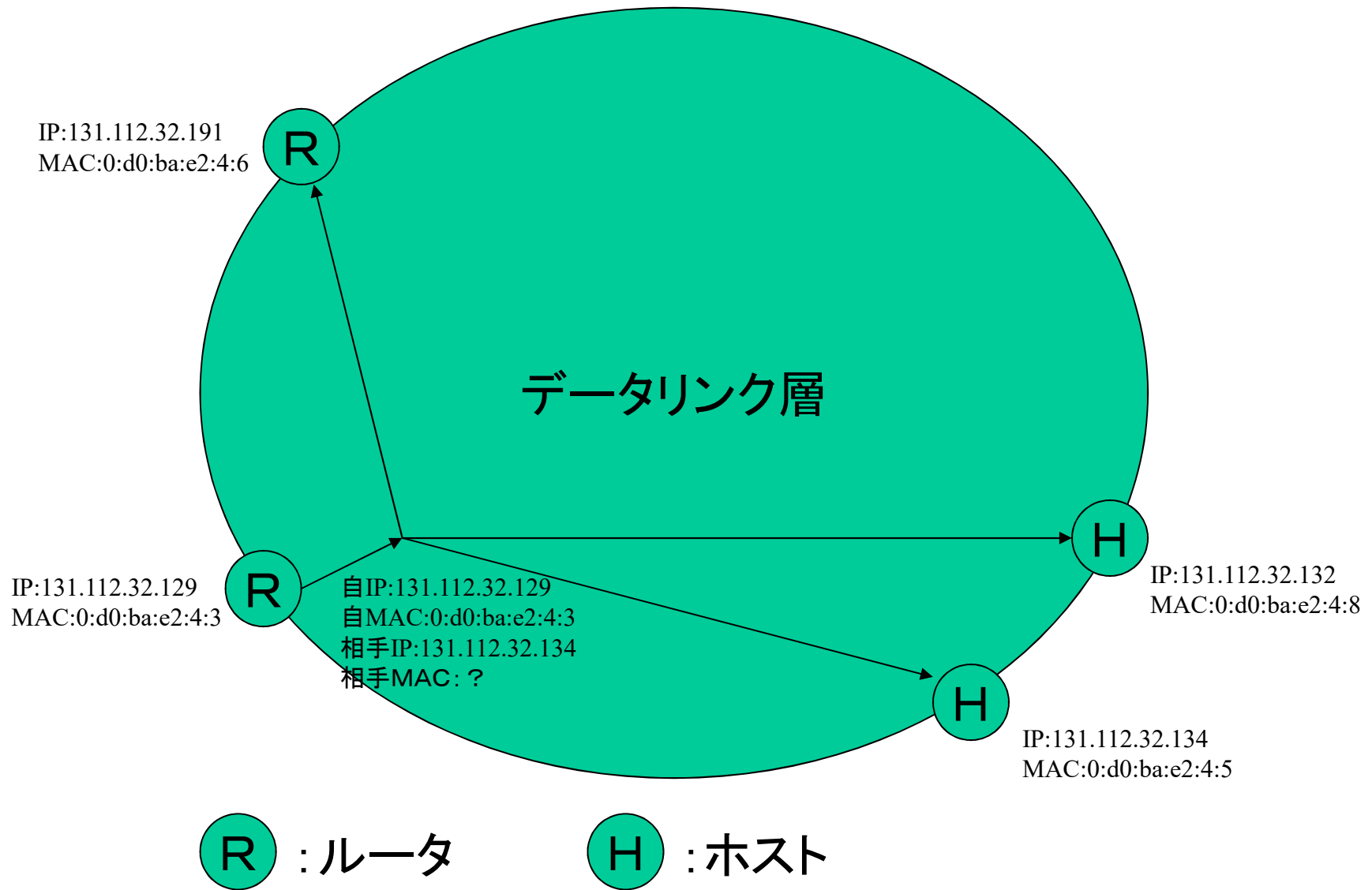
ICMP Time Exceeded

ルータによるパケットの中継 イーサネットとMACアドレス

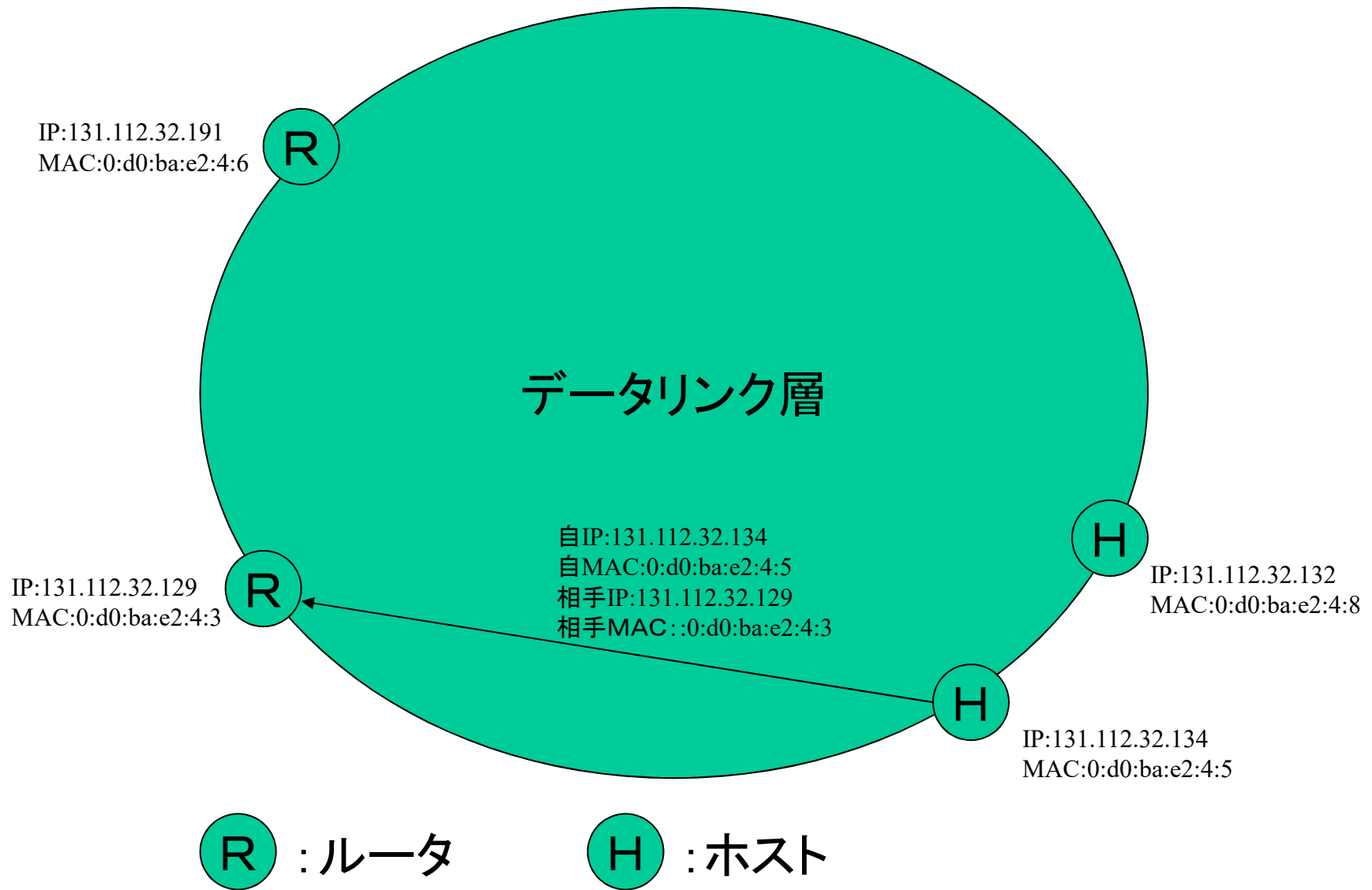
- 目的地がルータに直結したデータリンク
 - アドレス範囲(有効上位ビット)は静的設定済
 - 目的IPアドレスからMACアドレスを知り、中継
- 目的地が遠方のデータリンク
 - アドレスから、次段ルータを経路表で検索
 - 次段ルータのMACアドレスを知り、中継

ARP (Address Resolution Protocol、RFC826)

- IPアドレスとイーサネット(MAC)アドレスの対応付け
 - 同一データリンク内で、相手のIPアドレスがわかっているとき、そのMACアドレスを調べる
 - 相手のIPアドレス(と自分のIP、MACアドレス)を含むARP Queryをデータリンク層でブロードキャスト
 - 相手が答える
- IP、MACアドレスの重複を検出可能



ARP問い合わせ



ARP回答

まとめ

- IPv4ではネットワーク中ではほとんど何もやらない
 - 目的地への配送
 - フラグメンテーション
 - TTLの管理
- ICMPはIP層、トランスポート層を管理
- ARPによりMACアドレスは動的に取得