インターネットインフラ特論
５．IPv6、ND、新世代ネット

太田昌孝

mohta@necom830.hpcl.titech.ac.jp

ftp://chacha.hpcl.titech.ac.jp/infra5j.ppt

独占の得失

- 得
  - インターネットの成功、電話網、放送網の消滅
    - 情報通信の低価格化、高速化
- 失
  - IPv4の成功、IPv6の消滅？
    - ネットワークの破綻

IPv6は絶対に普及しない

- IPv6で動くアプリケーションはすべてIPv4でも動く
  - 生態学的地位が同一
- IPv4サービスは商売に直結している
  - IPv4には規模の利益
- IPv6は、民間の自由競争ベースでは、絶対に普及しない
  - 政策的誘導が必須

IPv6は絶対に必要（？）

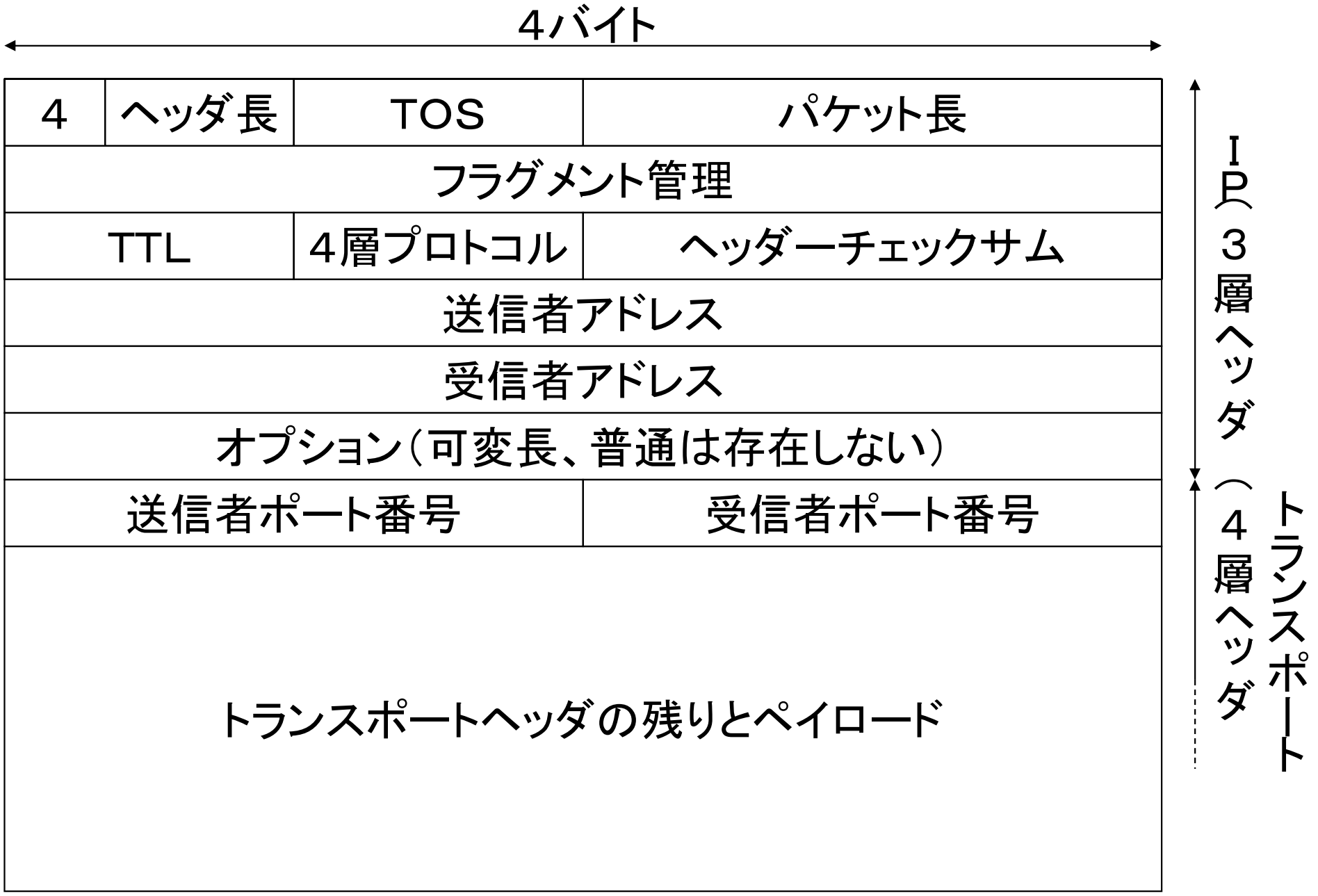
- IPv4では40億台の端末しかもてない
- IPv6では40億人がそれぞれ40億台の端末をもっても大丈夫
- 問題は、時期
  - IPv4アドレスが枯渇する以前
    - いつ枯渇するのか？
      - 現在は、IPv4アドレスの消費を厳しく抑える政策
      - 「後10年で枯渇？」と20年前から言われていたが、、、

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

IPv6 （Internet Protocol Version 6、RFC2460）

- 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
  - TTLの管理
  - （IPオプション）
  - （QoS保証？）



IPv4パケットフォーマット

IPv6パケットフォーマット

IPv6固有のヘッダ

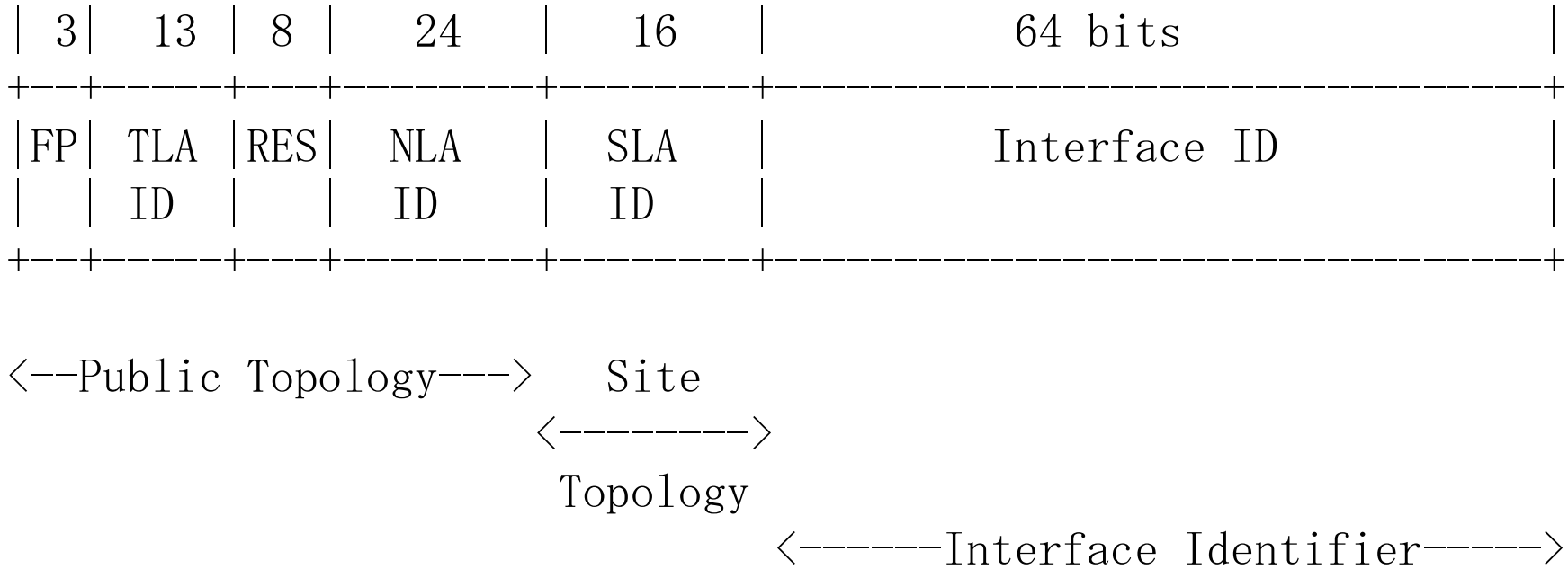
- フローラベル
  - ソースアドレスとフローラベルの対で、特定の通信を識別
    - QoS保証が楽に？！？
- 次ヘッダ
  - IPv4のオプション＋プロトコル
  - ヘッダをつなげて最後がトランスポートヘッダ
  - 次ヘッダさえなければQoS保証は元々楽

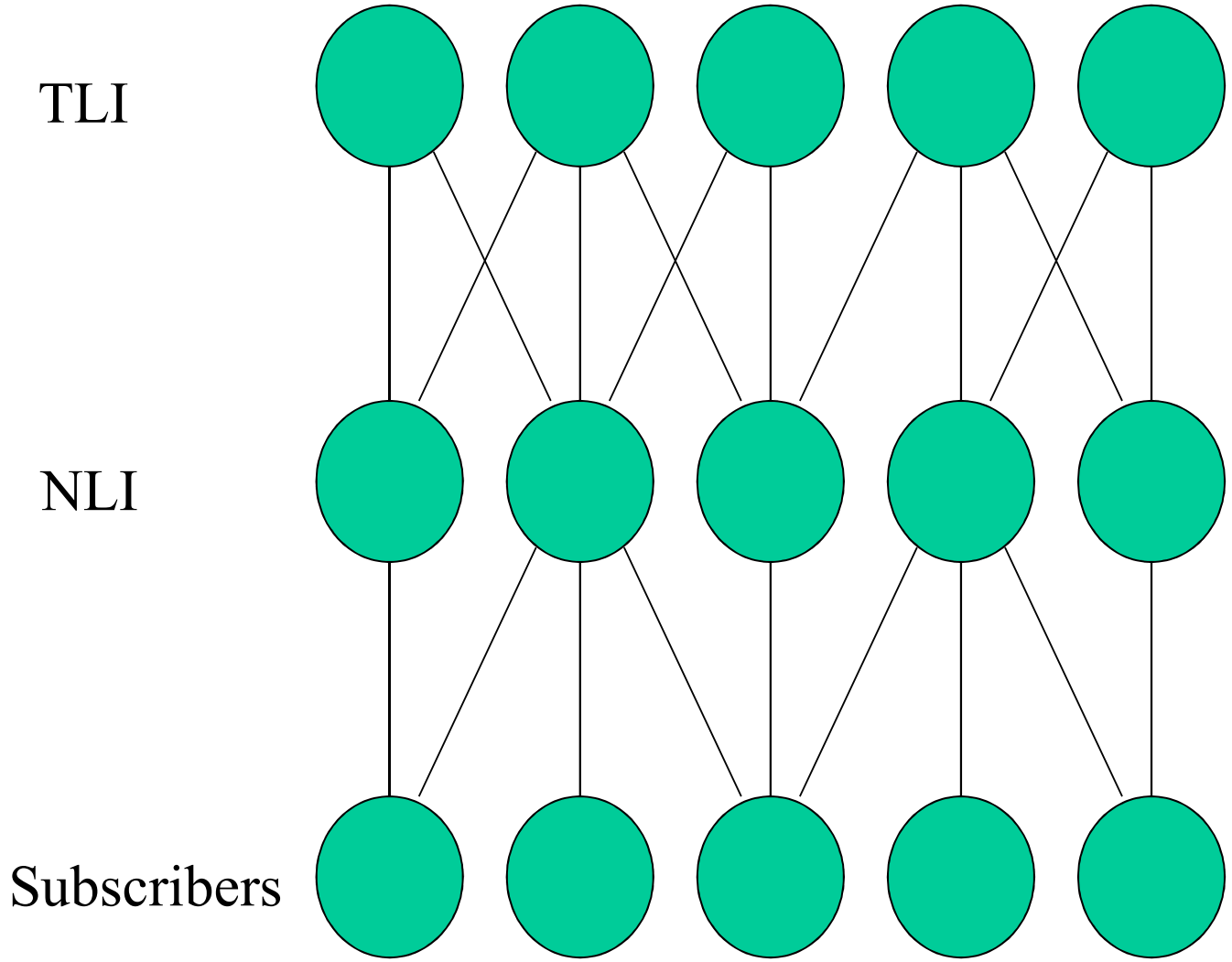
オプションIPヘッダを使わない（が使えない）わけ

- ルータで処理が必要なオプション
  - ルータの処理が複雑になる
    - ルータが遅くなる
    - ルータが落ちる
  - そもそも、ルータで処理が必要か？
- ルータで処理が不要なオプション
  - トランスポート層以上のオプション
- つまりはエンドツーエンド原理

IPv6アドレスの構造の初期提案（RFC2374）

- 強い階層構造
- ISPレベルで2段階
  - TLA （Top Level Aggregater）
  - NLA （Next Level Aggregater）
- 個別加入者（Subscriber）は65536個のサブネットをもてる
  - （Subscriber Level Aggregater）
- 各サブネット内は64ビットのアドレス


3
13
8
24
16
64 bits
FP
TLA ID
RES
NLA ID
SLA ID
Interface ID
<--Public Topology--->
Site
<-------->
Topology
<------Interface Identifier----->
IPv6アドレスの構造



Typical Scenario of IPv6 ISPs with Multihoming

IPv6が普及しないと

- IPv4アドレスは枯渇させてはならない
  - IPv4アドレス割当をますます抑制
  - IPアドレスがないと、インターネットは使えない
    - IPv4アドレスの節約技術（NAT）の発展
      - IPv4アドレスを複数の端末で使いまわす
        - » 個々の端末はインターネットに直接つながらない
        - » 従来のNATは、途中のゲートウェイで複雑な変換
        - » ゲートウェイを通るアプリケーションは限られる
  - NATによるインターネットの崩壊が進行中

IPv6の魅力？

- フラグメントなし
- PMTUD
- ND（Neighbor　Discovery）
  - コンフィギュレーション不要
  - 自動リナンバリング
- QoS保証
- セキュリティ
- モビリティ

IPv6のフラグメント

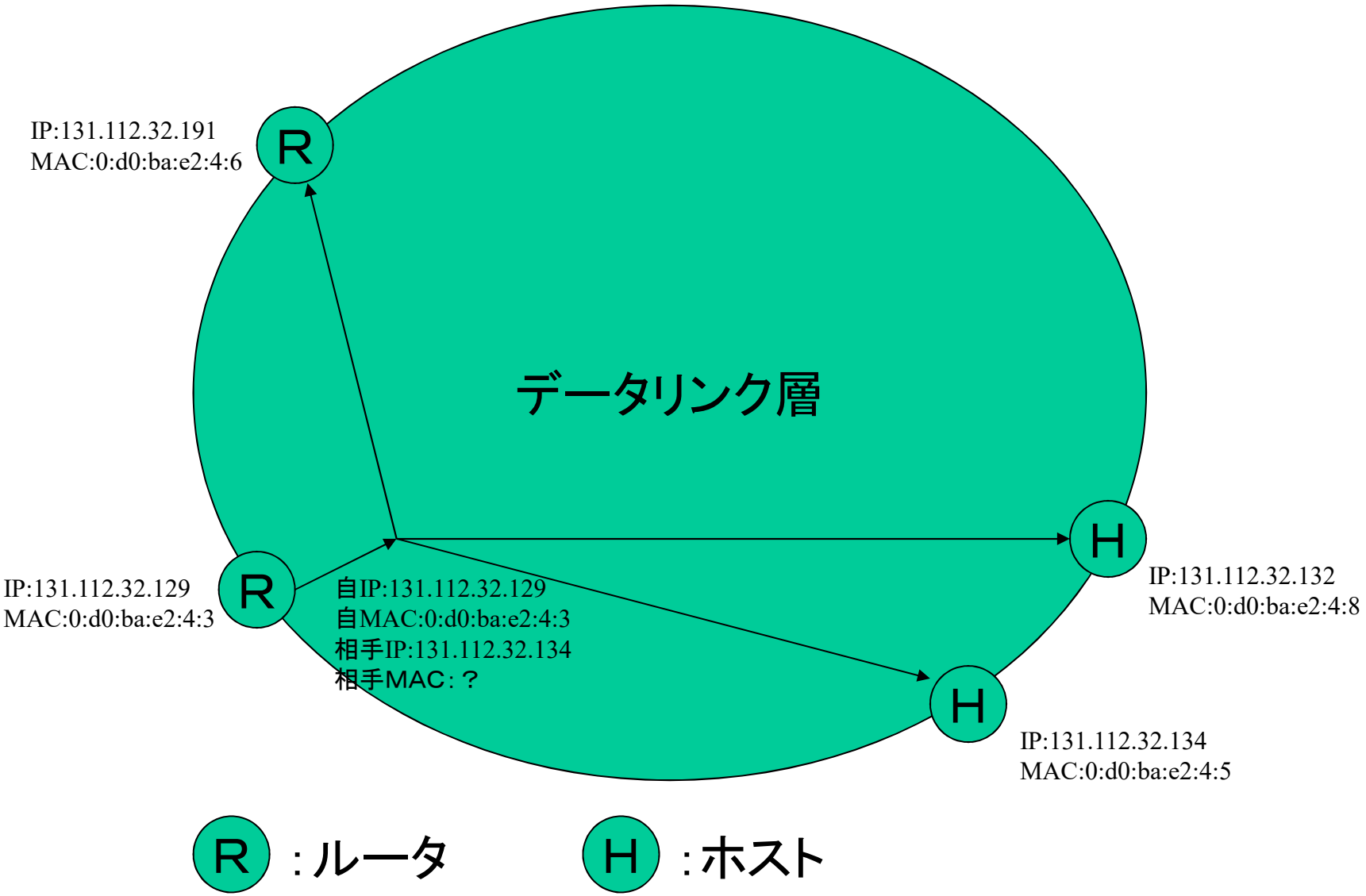
- 途中ルータでのフラグメントは禁止
  - 常にICMPエラーが帰る
  - 最初のホストでなら可
  - ルータの負荷を軽減？？？
- フラグメントを防ぐため
  - リンクのMTUは、最低1280バイト
    - イーサネットの1500バイトを意識
- PMTUDにより経路上の最少MTUを計測

Path MTU Discovery

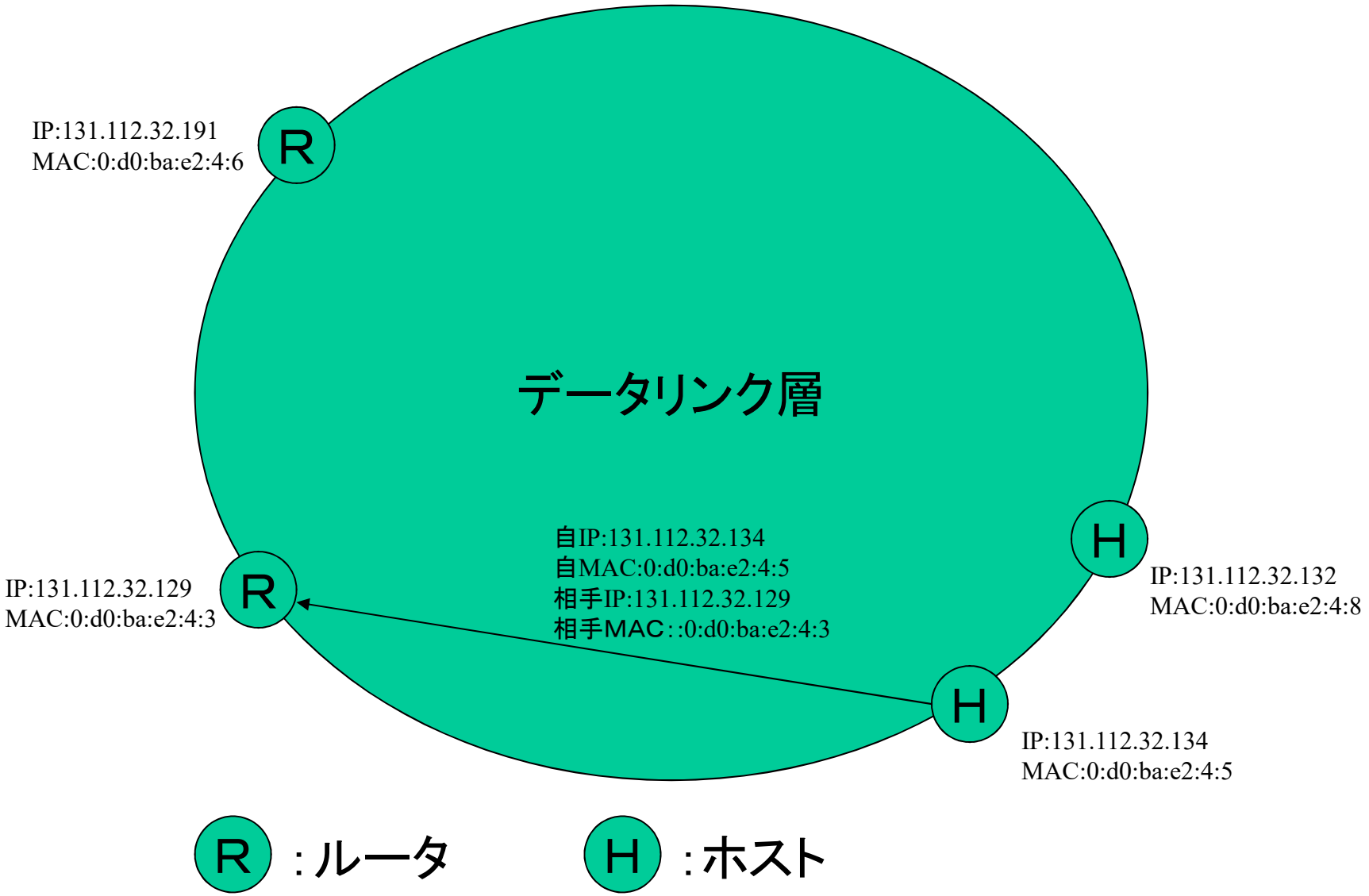
- 各トランスポートコネクションで
  - ある大きさのパケットをだしてみて
    - エラーがなければPMTUは丁度か大きい
      - 次にもうすこし大きなパケットを出す
    - エラーがあればPMTUは小さい
      - パケットの大きさを縮める
  - 経路は動的に変動するので定期的に再試行
    - ルータに定期的に負荷がかかる！！！
  - コネクションがなければ使えない
    - DNS、マルチキャスト等には適用不可

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


IP:131.112.32.191
MAC:0:d0:ba:e2:4:6
R
データリンク層
H
IP:131.112.32.132
MAC:0:d0:ba:e2:4:8
IP:131.112.32.129
MAC:0:d0:ba:e2:4:3
R
自IP:131.112.32.129
自MAC:0:d0:ba:e2:4:3
相手IP:131.112.32.134
相手MAC：？
H
IP:131.112.32.134
MAC:0:d0:ba:e2:4:5
R ：ルータ
H ：ホスト

ARP問い合わせ


IP:131.112.32.191
MAC:0:d0:ba:e2:4:6
R
データリンク層
自IP:131.112.32.134
自MAC:0:d0:ba:e2:4:5
相手IP:131.112.32.129
相手MAC：:0:d0:ba:e2:4:3
H
IP:131.112.32.132
MAC:0:d0:ba:e2:4:8
IP:131.112.32.129
MAC:0:d0:ba:e2:4:3
R
H
IP:131.112.32.134
MAC:0:d0:ba:e2:4:5
R ：ルータ
H ：ホスト

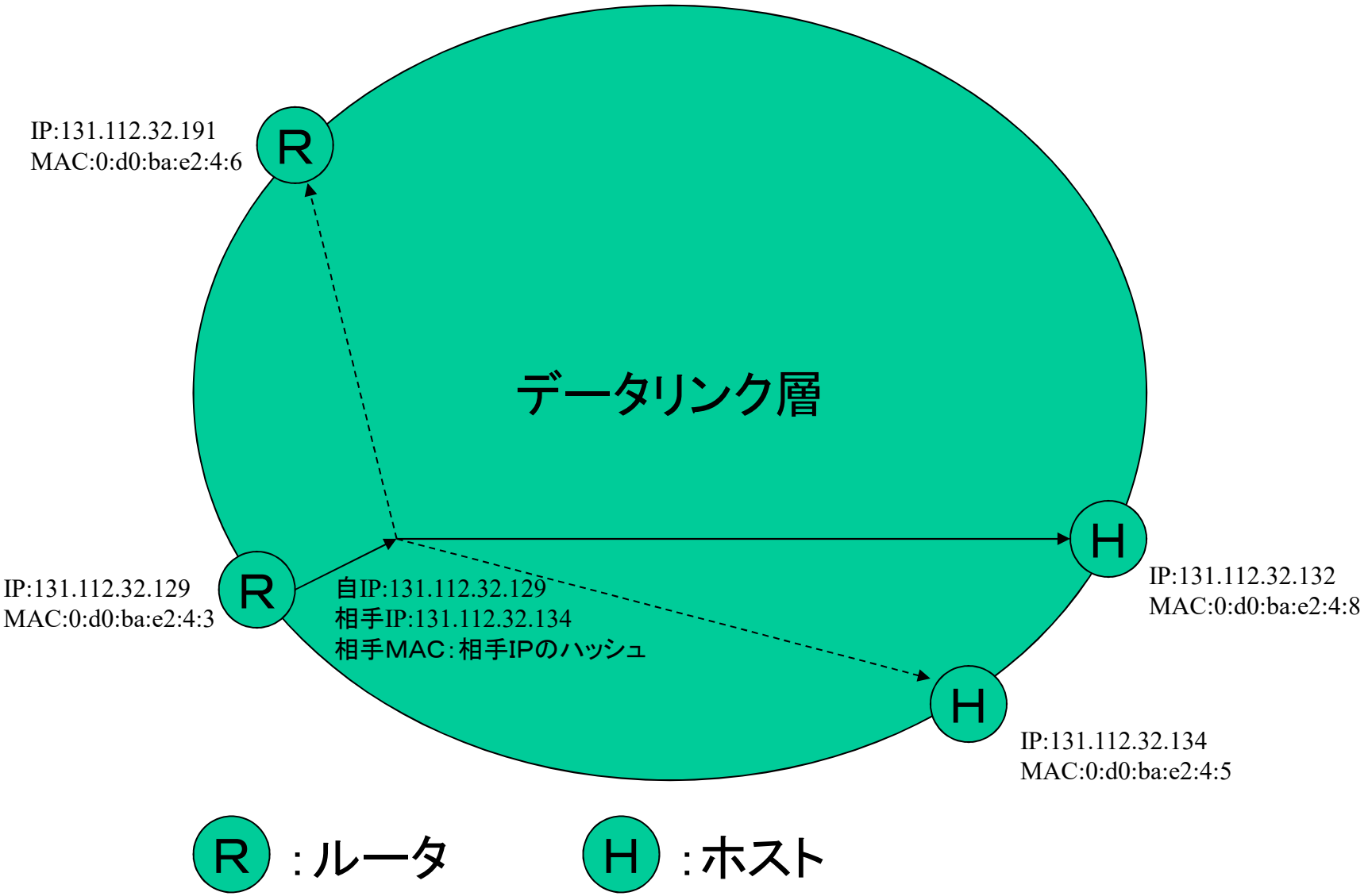
ARP回答

ARPの問題

- ARPがかえってくるまでもとのパケットを保持しなければならない
- ブロードキャストを使う
  - 大きなリンクでは使えない？
    - そもそもデータリンク層がポイントツーポイントならARPは不要

Neighbor Discovery（RFC2461）

- リンクマルチキャストを多用
  - 当初案では、パケットはいきなりリンクマルチキャストアドレスになげる（結局とりさげ）
    - 正しい相手のみが受け取る可能性が大きい
    - リンクマルチキャストがちゃんと動かないと、、、
- ホストとルータを峻別
  - ホストの構成を自動化？
  - ホストのリナンバリングが簡単？
  - 実はエンドツーエンド原理違反



NDによるパケット送出（当初案）

コンフィギュレーション不要 (Stateless Autoconfiguration)

- MAC（マッキントッシュ、当時）の真似
  - 家庭に何台も入るのに、MAC程度にしたい
  - アドレスが16Bになったのは、これが理由？
- 無理
  - セキュリティ0
    - 閉じたLANならいいが
  - DNSへの名前の登録すらできない
    - インターネットでは相手にされない

自動リナンバリング

- IPv6ではCIDRが当然
  - ISPをかえると、アドレスがかわる
- アドレスを変えると
  - DNSの内容も自動的に更新したい
- 無理
  - DNSにはDNSサーバのアドレスは生で入っている
    - DNSを深く理解した人がうまくやれば、なんとか

QoS保証

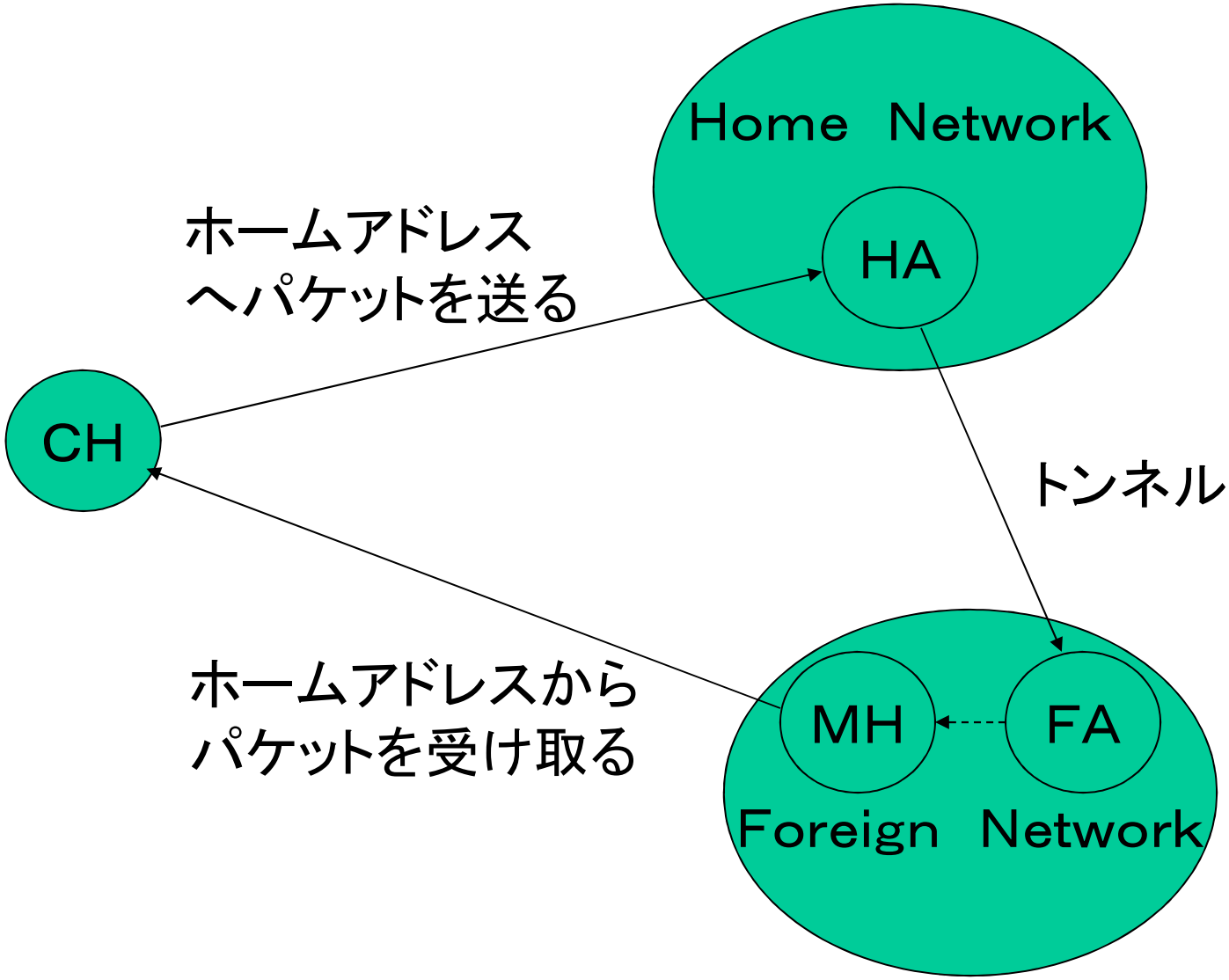
- フローラベルによるQoS保証
- そもそもRSVP動くの？
- フローラベルはどう使うの？
- フローラベルなしだと、IPv4のほうが楽

セキュリティ

- IPv6ではIPSECを標準実装、、、
- セキュリティには鍵が必要
  - IPSECだけでは使えない

モビリティ

- IPv6ではモビリティを標準実装
  - モバイルホストだけでなく、その固定ホストもモバイル対応
    - 固定ホストと協力して三角形を解消可能？
- IPv6のは、やたら複雑
- IPv4のモビリティは、普通に動作する



三角形型パケットのやりとり

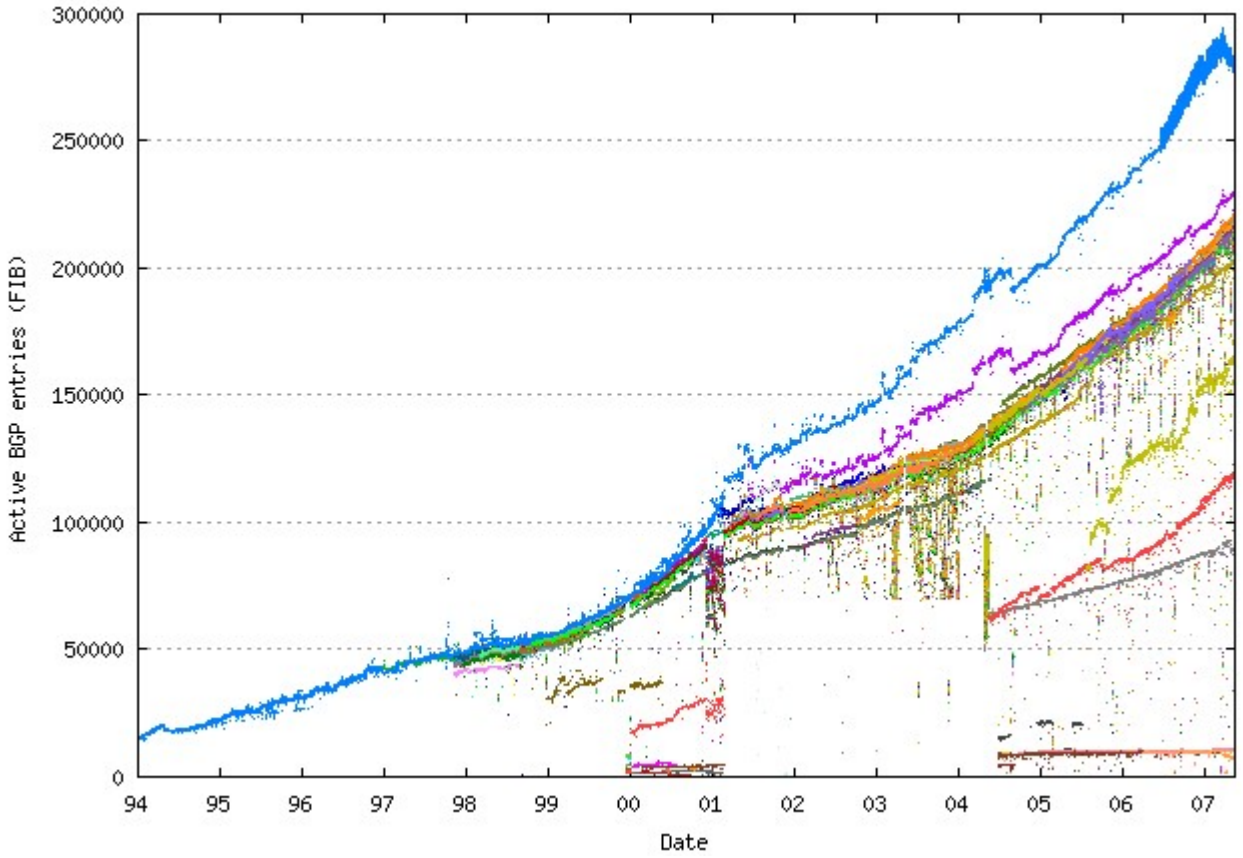
IPv6のほんとの意義

- アドレス空間が大きくなる
- ルーティングテーブルが小さくなる
  - マルチホーミングをきちんと解決すれば

IPv4 Routing Table Size

http://bgp.potaroo.net/

Growth of the BGP Table - 1994 to Present



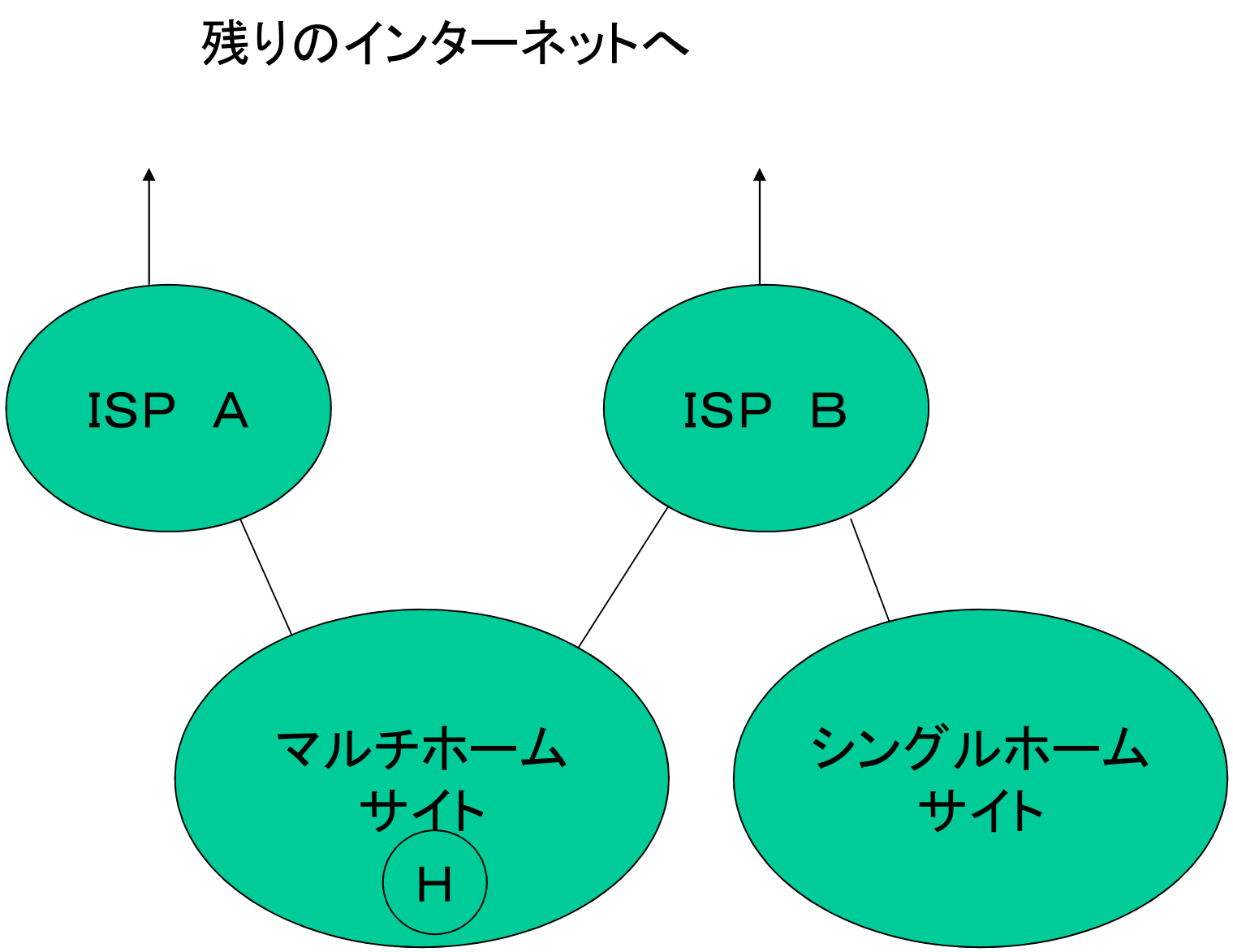
(Data Gathered from AS1221 and Route-Views)

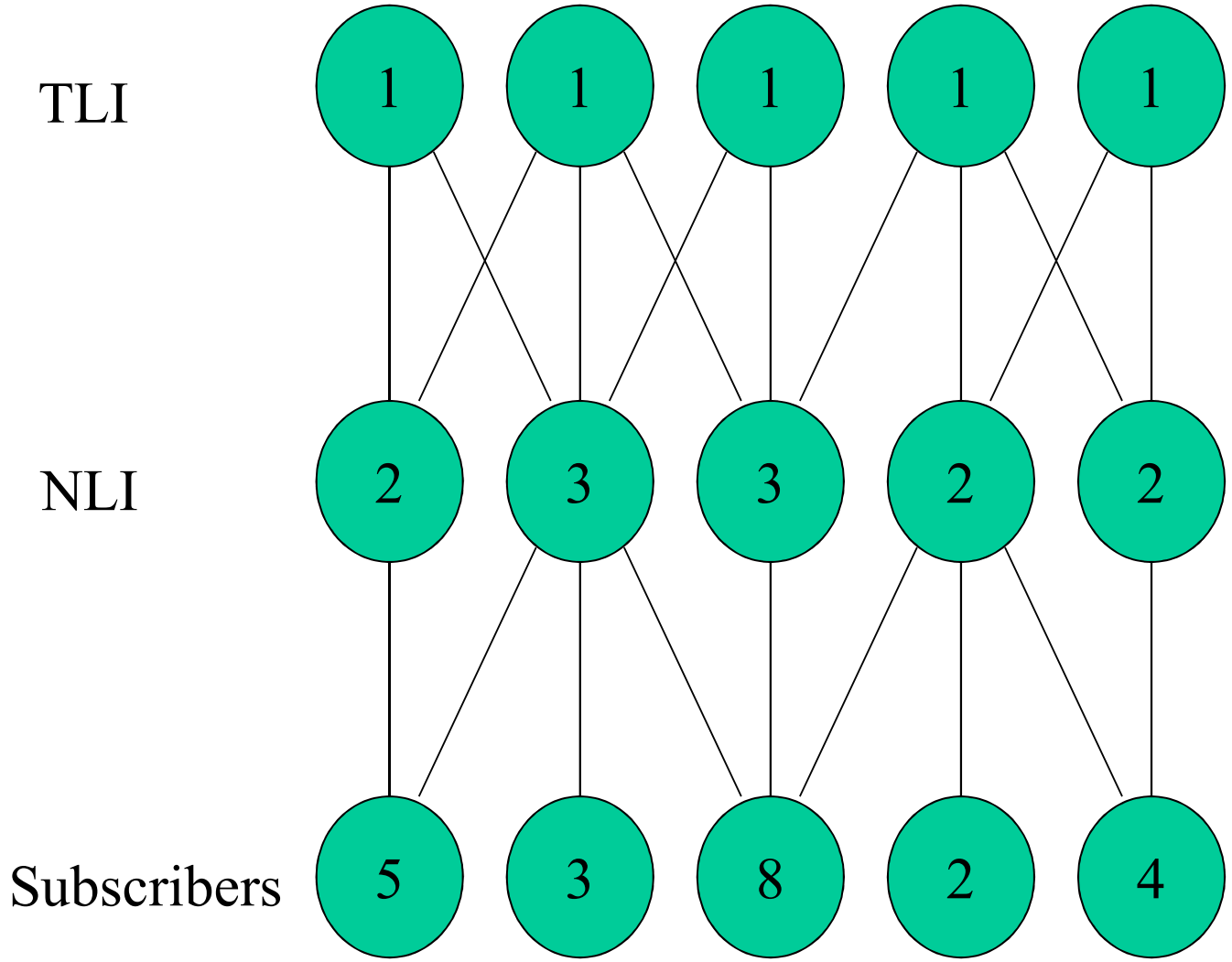
経路の縮約が不可能な場合

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

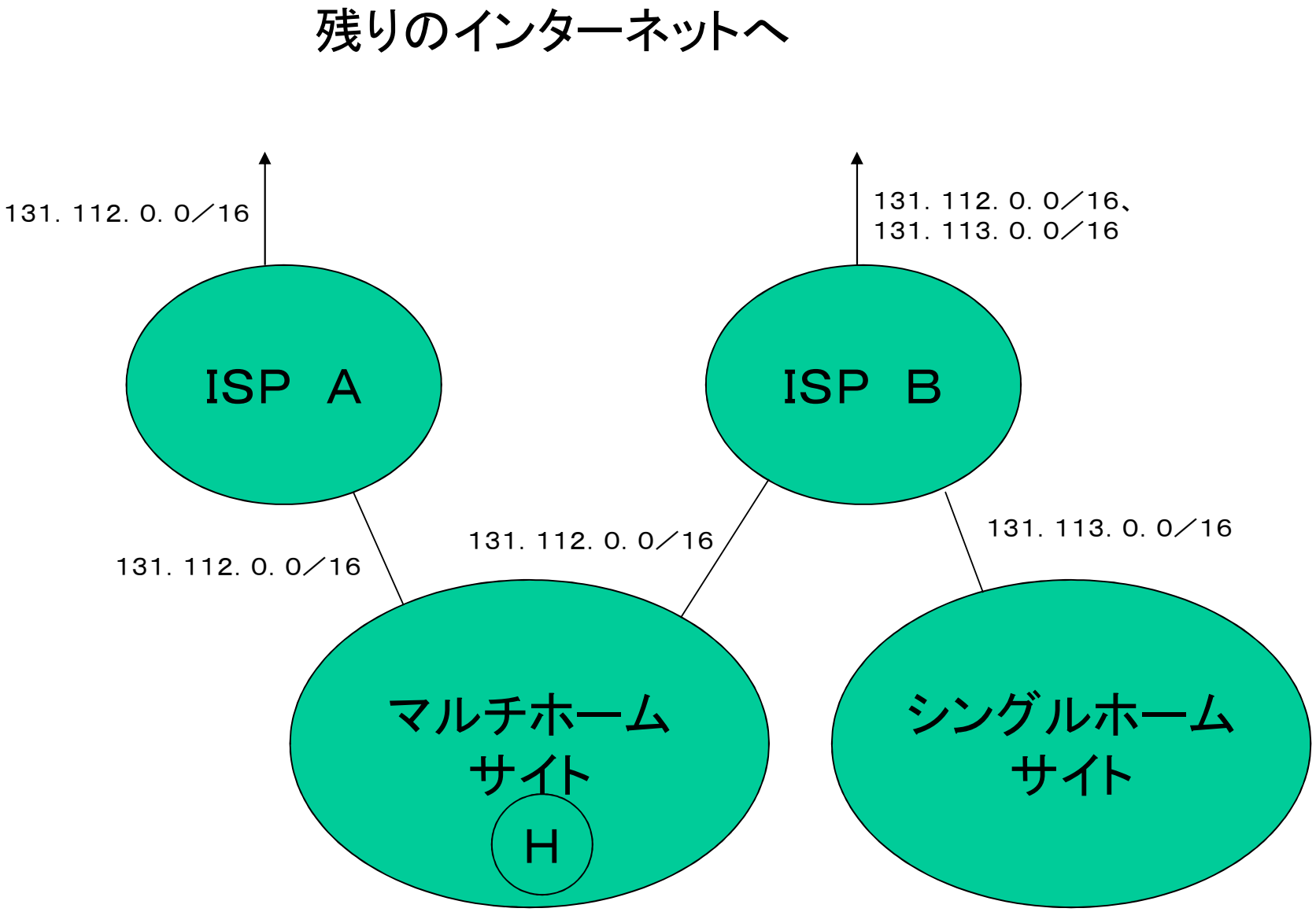
マルチホーミング

- 複数の上流ISPをもつ
  - どちらかがこけても大丈夫
- 信頼性のあるサービス（含ISP）には必須
  - IPv6ではNLISPは複数のTLISPに接続したいが、、、
- ルーティングによるマルチホーミングではISPはTLAしかもてない


残りのインターネットへ
ISP A
ISP B
マルチホーム
サイト
H
シングルホーム
サイト

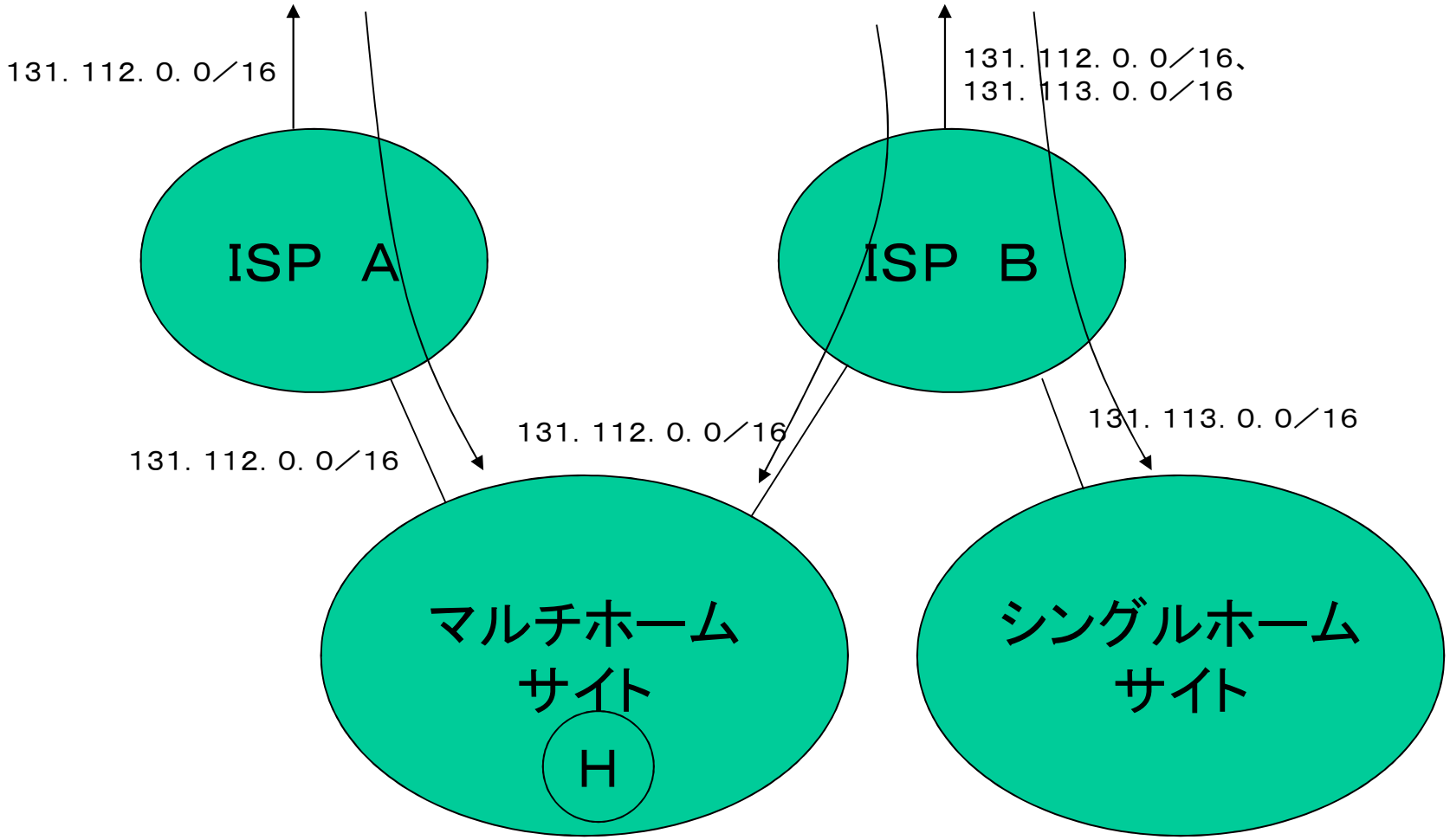


Number of Prefixes with E2E Multihoming

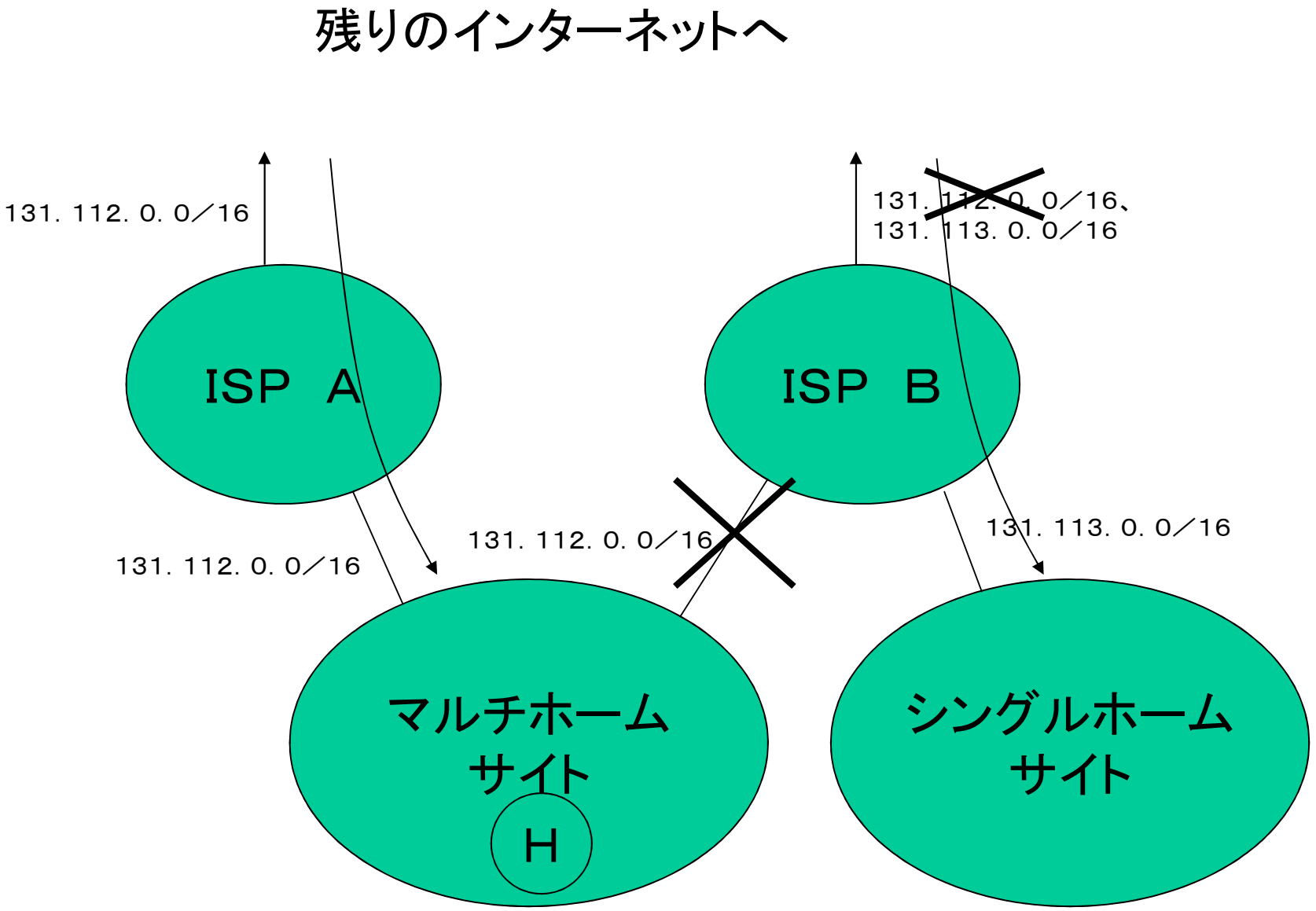


ルーティングによるマルチホーミング

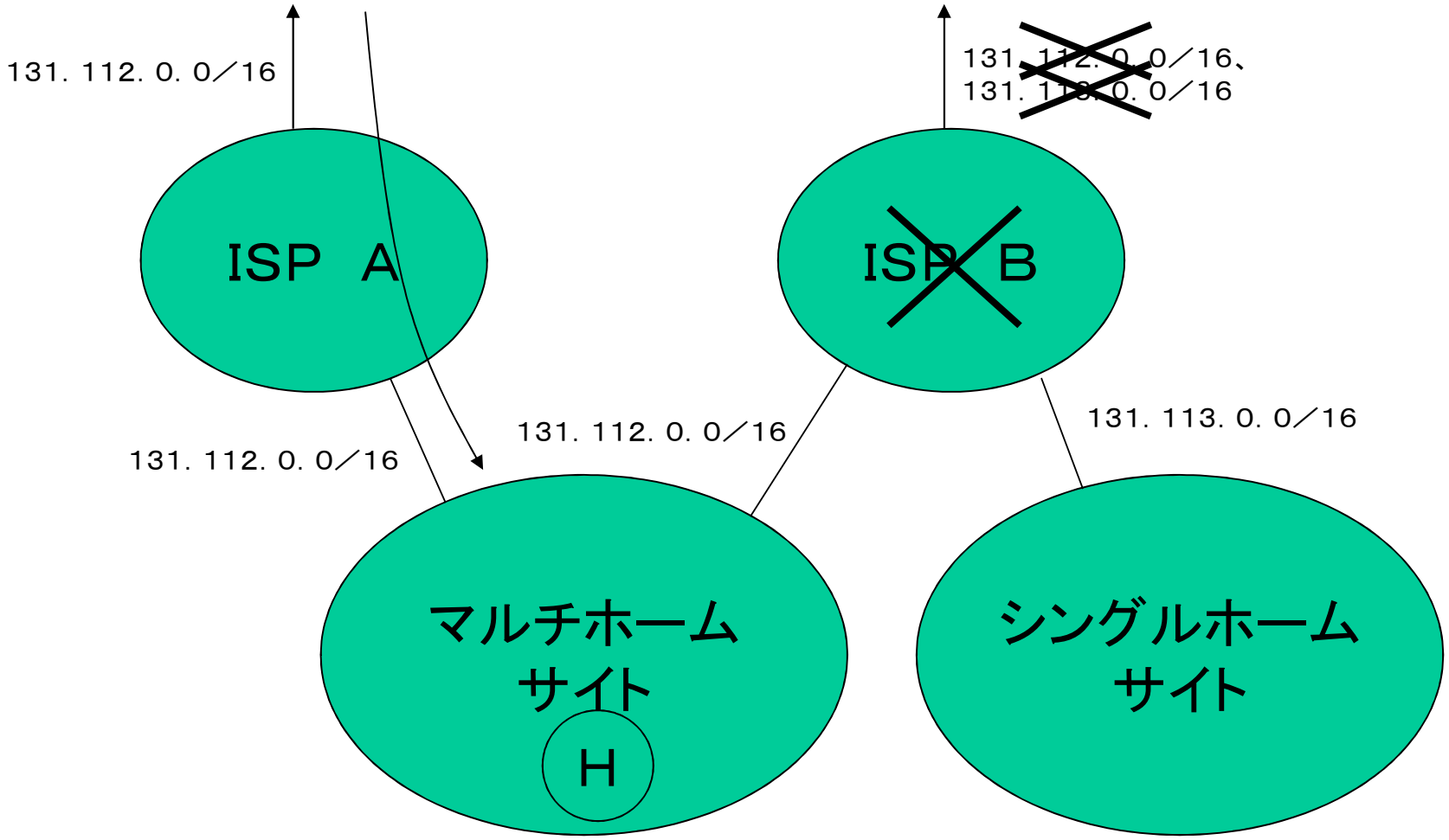
残りのインターネットへ



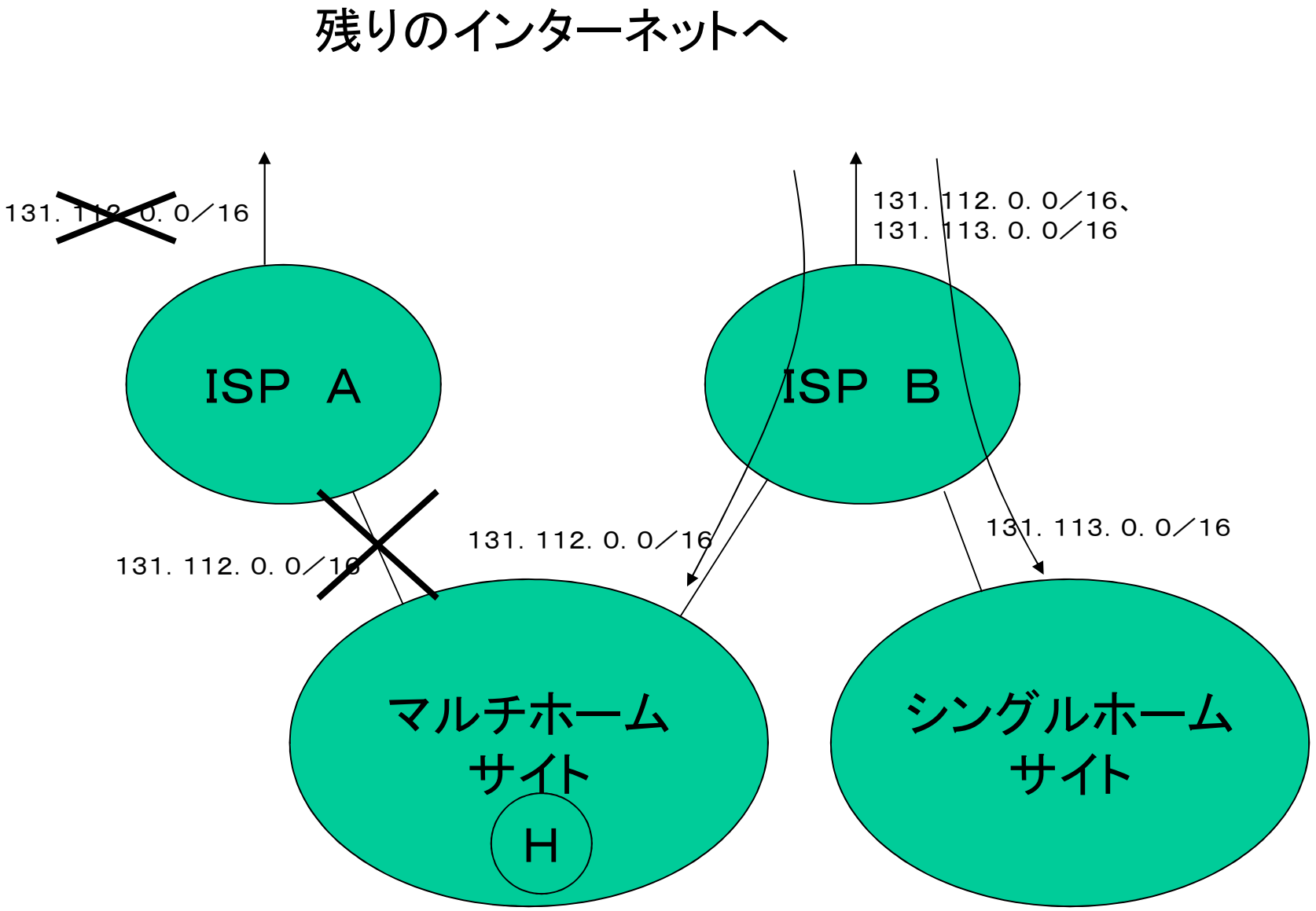
ルーティングによるマルチホーミング


残りのインターネットへ
131. 112. 0. 0／16
131. 112. 0. 0／16、
131. 113. 0. 0／16
ISP A
ISP B
131. 112. 0. 0／16
131. 112. 0. 0／16
131. 113. 0. 0／16
マルチホーム
サイト
H
シングルホーム
サイト
ルーティングによるマルチホーミング

残りのインターネットへ

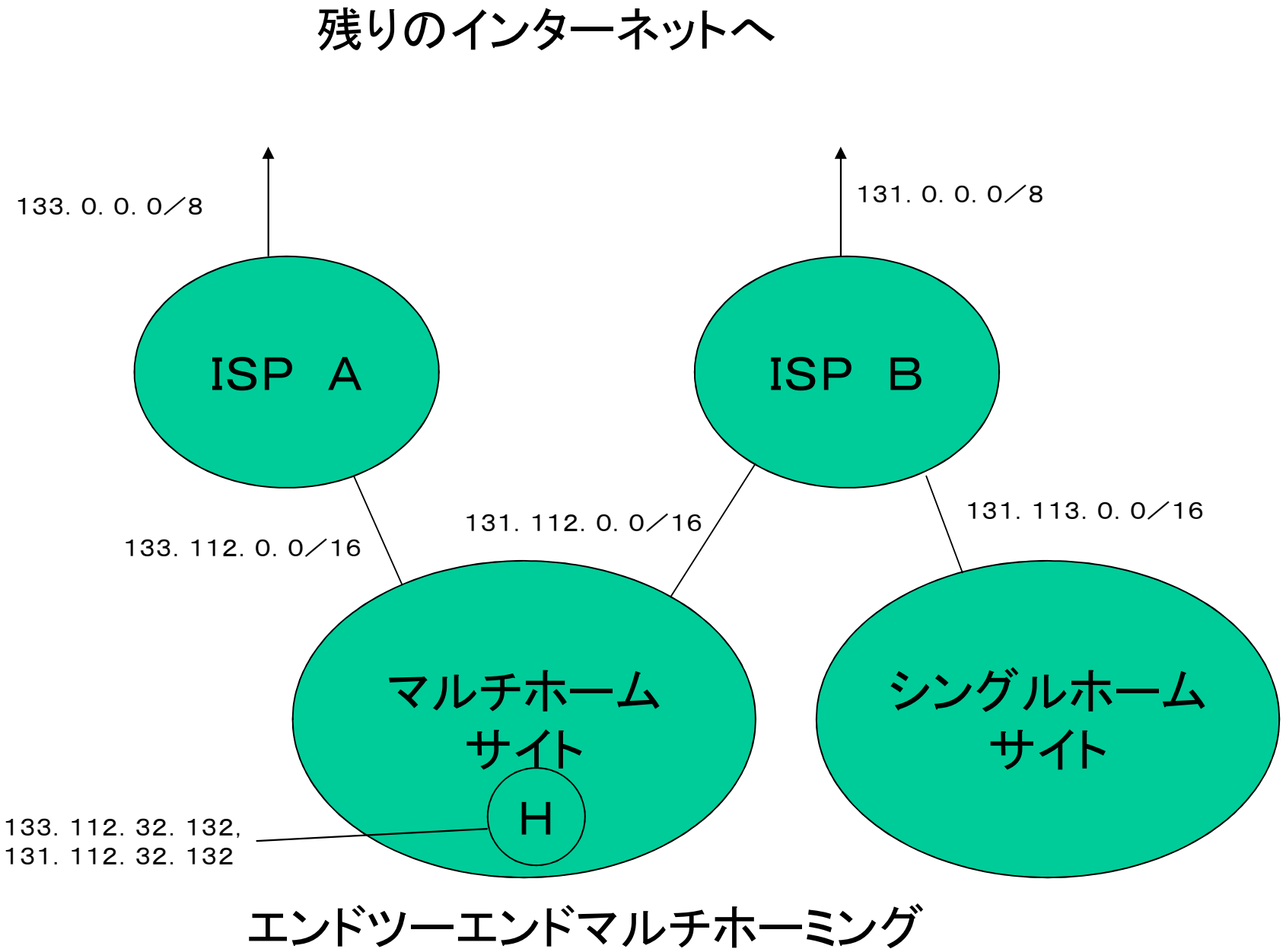


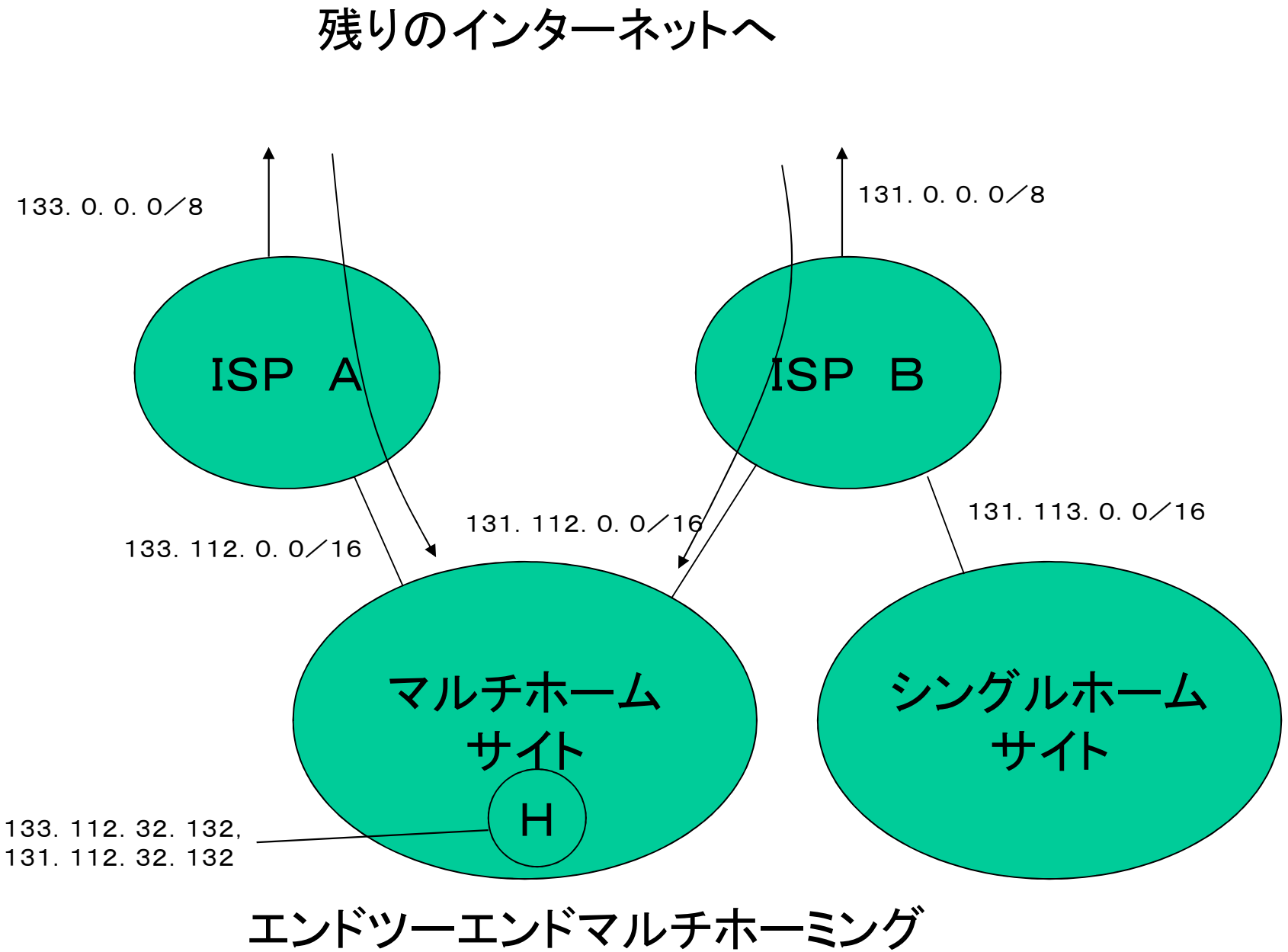
ルーティングによるマルチホーミング

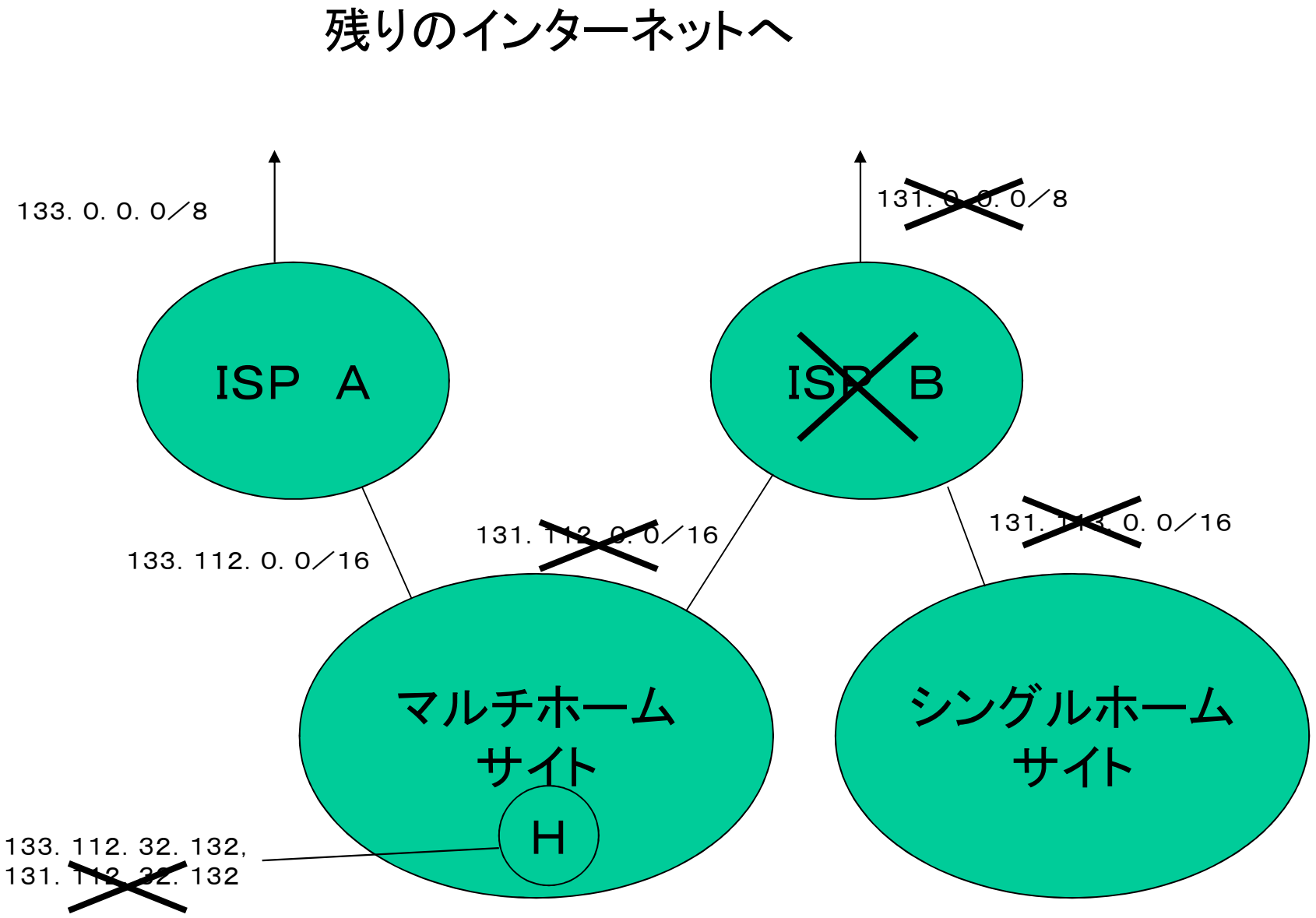

残りのインターネットへ
131. 112. 0. 0／16
131. 112. 0. 0／16、
131. 113. 0. 0／16
ISP A
ISP B
131. 112. 0. 0／16
131. 112. 0. 0／16
131. 113. 0. 0／16
マルチホーム
サイト
H
シングルホーム
サイト
ルーティングによるマルチホーミング

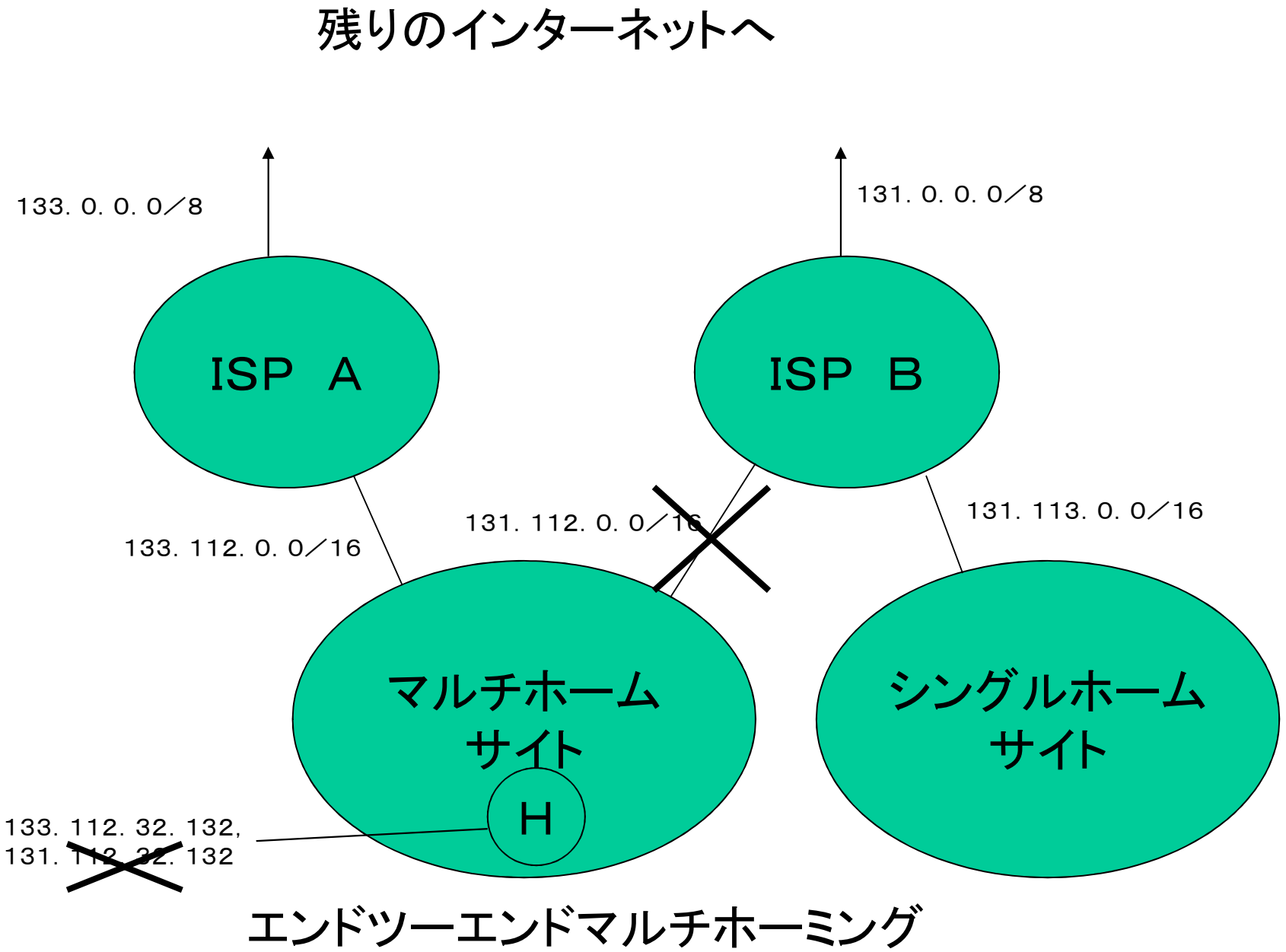
エンドツーエンド
マルチホーミング

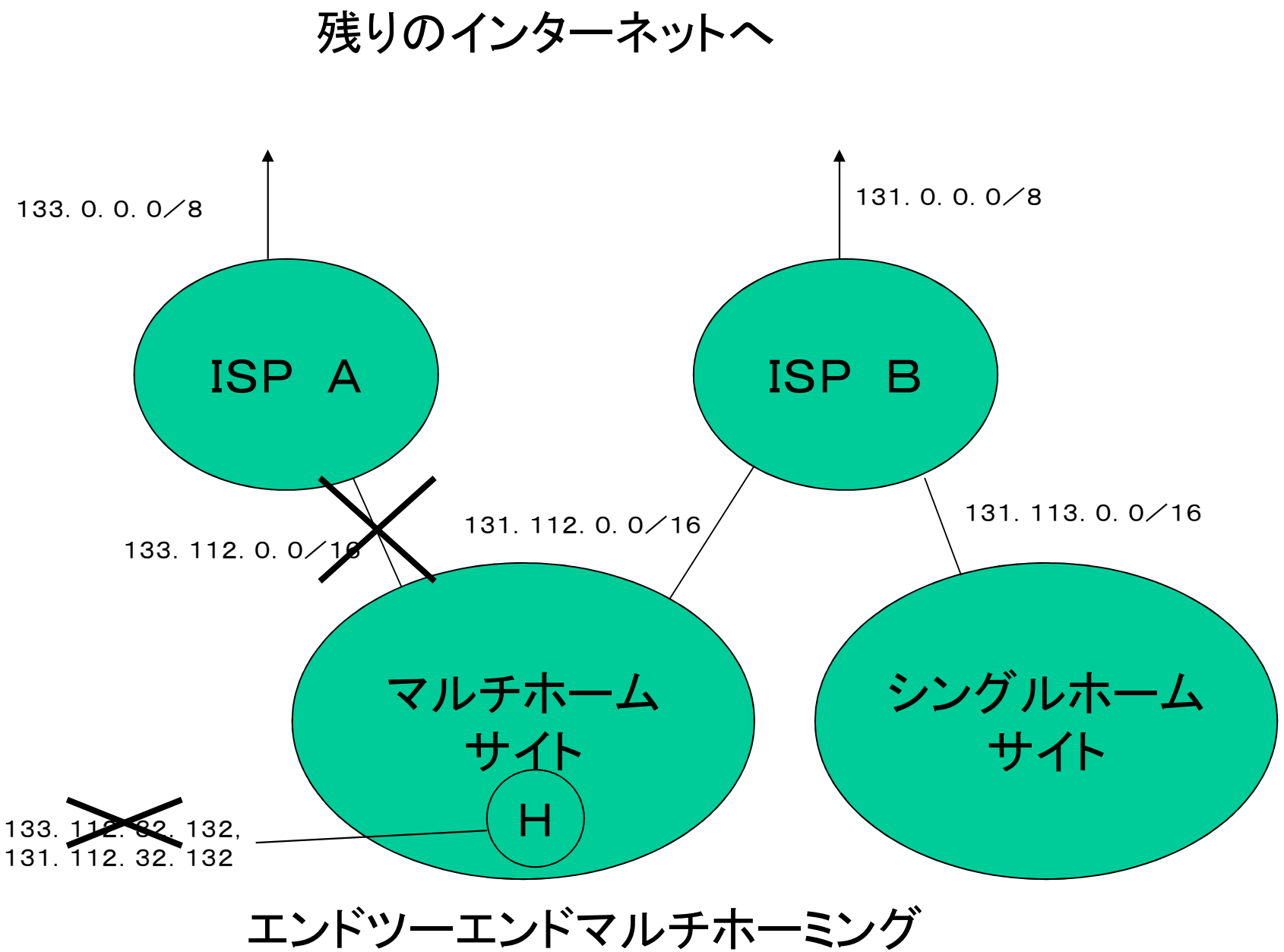
- ホストは複数のアドレスをもつ
- ホストの通信相手は複数のアドレスを自分で試す
  - 通信不能は、まず大域的経路表でチェック
  - どれかでつながれば通信は成立
  - 通信中にタイムアウトなどがおきれば、他のアドレスを試す
- ルーティングマルチホーミングは不要


残りのインターネットへ
133. 0. 0. 0／8
131. 0. 0. 0／8
ISP A
ISP B
133. 112. 0. 0／16
131. 112. 0. 0／16
131. 113. 0. 0／16
マルチホーム
サイト
H
シングルホーム
サイト
133. 112. 32. 132,
131. 112. 32. 132
エンドツーエンドマルチホーミング


残りのインターネットへ
133. 0. 0. 0／8
131. 0. 0. 0／8
ISP A
ISP B
133. 112. 0. 0／16
131. 112. 0. 0／16
131. 113. 0. 0／16
マルチホーム
サイト
H
シングルホーム
サイト
133. 112. 32. 132,
131. 112. 32. 132
エンドツーエンドマルチホーミング


残りのインターネットへ
133. 0. 0. 0／8
131. 0. 0. 0／8
ISP A
ISP B
133. 112. 0. 0／16
131. 112. 0. 0／16
131. 113. 0. 0／16
マルチホーム
サイト
H
シングルホーム
サイト
133. 112. 32. 132,
131. 112. 32. 132
エンドツーエンドマルチホーミング


残りのインターネットへ
133. 0. 0. 0／8
131. 0. 0. 0／8
ISP A
ISP B
133. 112. 0. 0／16
131. 112. 0. 0／16
131. 113. 0. 0／16
マルチホーム
サイト
H
シングルホーム
サイト
133. 112. 32. 132,
131. 112. 32. 132
エンドツーエンドマルチホーミング


残りのインターネットへ
133. 0. 0. 0／8
131. 0. 0. 0／8
ISP A
ISP B
133. 112. 0. 0／16
131. 112. 0. 0／16
131. 113. 0. 0／16
マルチホーム
サイト
シングルホーム
サイト
H
133. 112. 32. 132,
131. 112. 32. 132
エンドツーエンドマルチホーミング

IDとロケータの分離

- （IPv6）アドレス（16B）を、ID（8B）とロケータ（8B）に分離
  - IDは、ホストをグローバルに特定
    - 固定長のドメイン名のようなもの
  - ロケータは、ホストを含むサブネットへの経路
    - TL、NL等に階層化される
- IPv6では、結局採用されず
  - なら、アドレスは8Bでよかった？

ID・ロケータ分離の意義

- エンドツーエンドMHに便利
  - データ量が約半分に
- ロケータは途中で書き換え可能
  - ソースロケータの成りすまし防止
  - MIPでのトネリングが不要に

IPv4で、今なにが問題か？

- ルート情報の増大
  - 極めて深刻な問題
    - ルーティングテーブル
    - ルート計算
- アドレス空間の枯渇
  - いちおう問題

これまでの対策

- ルート情報の増大
  - CIDR
  - IPv6アドレッシングアーキテクチャ
- アドレス空間の枯渇
  - １２８ビットアドレス空間（IPv6）
    - じつは64ビット（＋α）
- 共有
  - Usage　Based　Allocation

結果

- Usage Based Allocation
  - アドレスをなかなか割り当てない
  - CIDR環境では新規割り当てはRenumberingを伴う
  - Renumberingは受け入れられない
  - その代替策としてのNATの発達
  - DHCP／PPPによるアドレス割り当ての発達
  - アドレス空間の枯渇は起こりにくい

インターネットの崩壊

- Renumberingは受け入れられない
  - DHCP／PPPによるアドレス割り当ての発達
    - 常時接続性の喪失
      - 少数の常時接続サーバによるプロキシー
      - ダイアルアップクライアント
        - » End to End原理の崩壊
- NATの発達
  - Global Connectivity原理の崩壊

悪循環（現実）

- NATがはびこるとIPv4アドレス空間はぎりぎり枯渇しない
  - IPv4アドレス空間がぎりぎり枯渇しないならIPv6は普及しない
    - IPv6が普及しないので、ぎりぎりのIPv4のアドレス空間割り当ては慎重になる
      - IPv4アドレスが割り当てられないと、NATがはびこる
- NATはインターネットを崩壊させる

好循環（理想）

- IPv6さえ普及するなら、アドレス空間はどしどしわりあててよい
  - IPv4アドレスをどんどん割り当てるとNATはなくなる
    - NATがなくなれば、IPv4アドレスは枯渇する
      - IPv4アドレスが枯渇すれば、IPv6に移行
- NATがなければ、Global　Connectivityは回復する

IPv6普及のために必要だった政策

- IPv4アドレス空間は「大量に」割り当てる
  - ただし、以下の条件を満たすISPに
    - End to End原理の遵守
    - Global Connectivity原理の遵守
  - 具体的には
    - IPv6への移行
    - 加入者へ複数のグローバルアドレスの割り当て
    - 24／7の超高速常時接続
    - グローバルアドレスのDNSへの登録

特定のISPを優遇するか？

- そんなことはしない
- 条件を満たすISPは同列に扱う
- 条件をそれなりにきびしくしないと、アドレス空間の奪い合い
  - 結果的に「大量の」割り当てにはならないので計画は失敗
- 条件をどこまで厳しくできるか、事前に有望そうなISPと打ち合わせる

具体的条件（１）

- IPv6への移行
  - 当面実験的サービスで十分
    - サービスの質などは不問
- 加入者へ複数のグローバルアドレスの割り当て
  - IPv4で8（＊２？）程度？
  - IPv6では2^80が当然
  - リナンバリングは不要に

具体的条件（２）

- ２４／７の超高速常時接続
  - アドレス動的割り当てでは常時接続の意味がない
- グローバルアドレスのＤＮＳへの登録
  - アドレス動的割り当てでは事実上不可能
- その他？

まとめ

- IPv6は必用だった
- IPv6は、現状では使えない
  - プロトコルの目的の見直しが必用
  - プロトコルの単純化が必用
  - 運用方法の見直しも必用
- 米国中心状態からの巻き返しのチャンスではある
- 今後、IPv6は普及するのか？

Inapplicability of Neighbor Discovery over Wireless LAN

Masataka Ohta

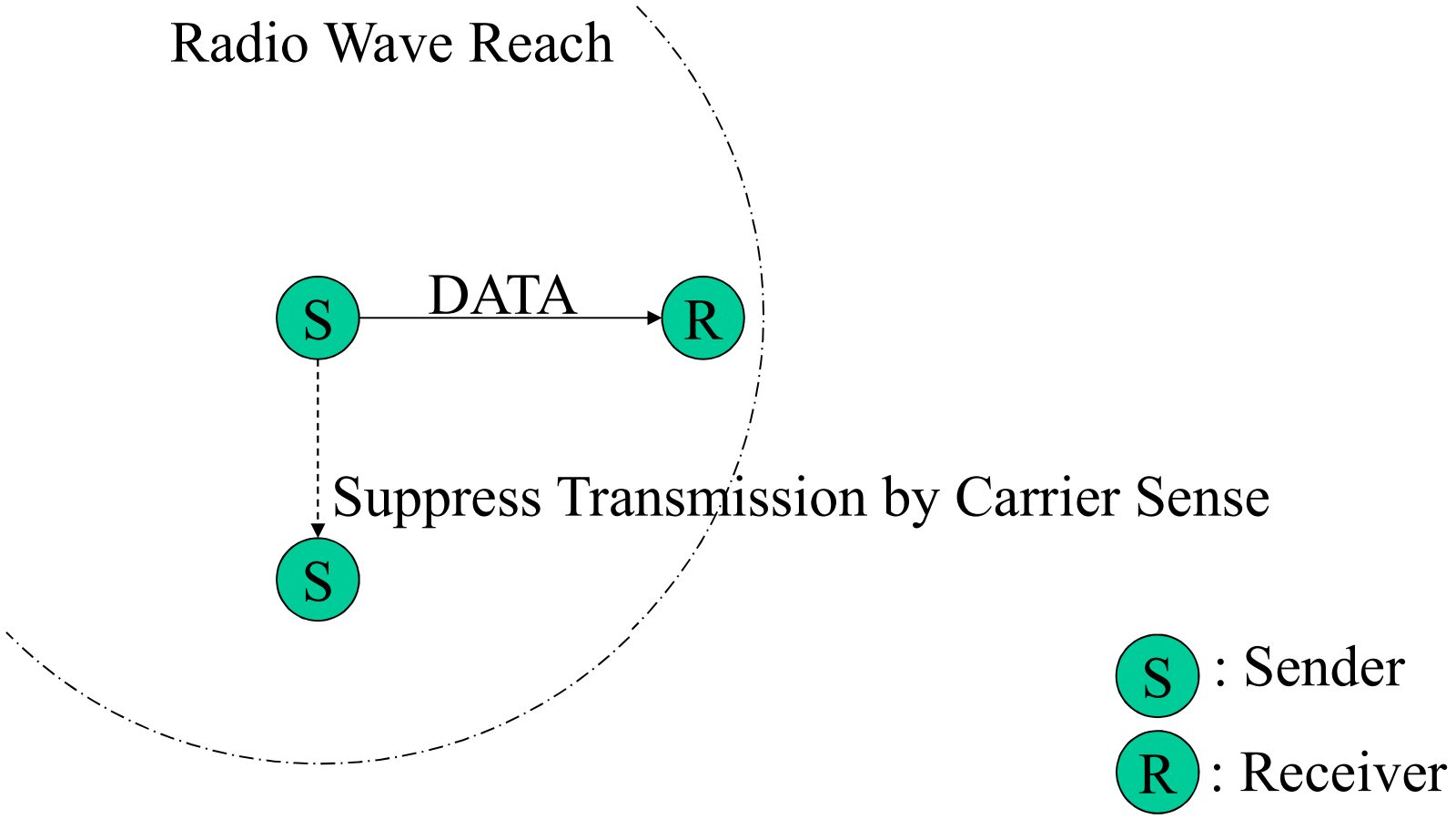
Tokyo Institute of Technology

mohta@necom830.hpcl.titech.ac.jp

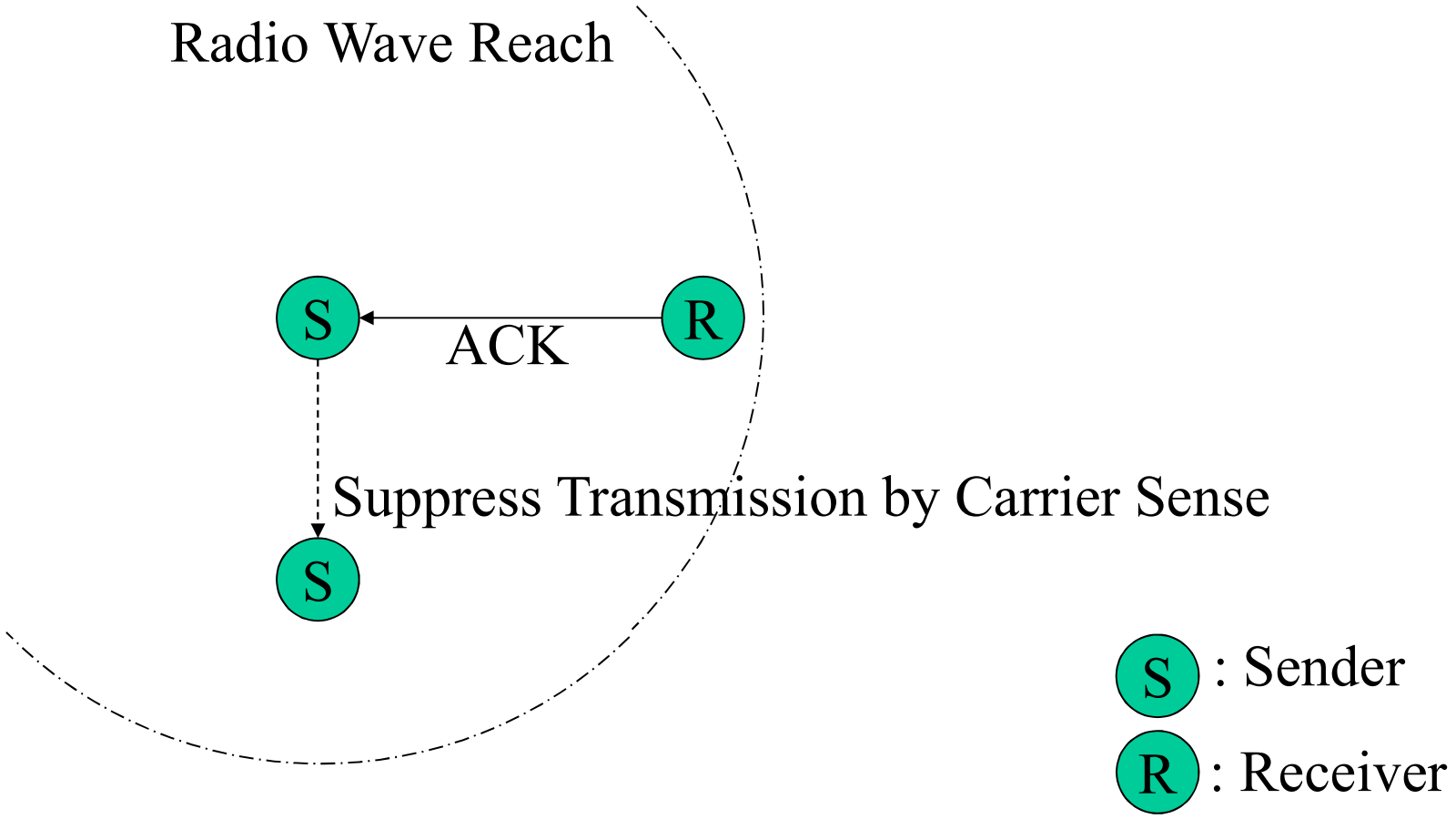
Wireless LAN of IEEE 802.11

- Relies on CSMA/CA
  - because of undetectable collisions
    - ACKs are the MUST for reliable communication
      - or packets are lost upon collisions
    - Unicast packets are ACKed and delivered reliably

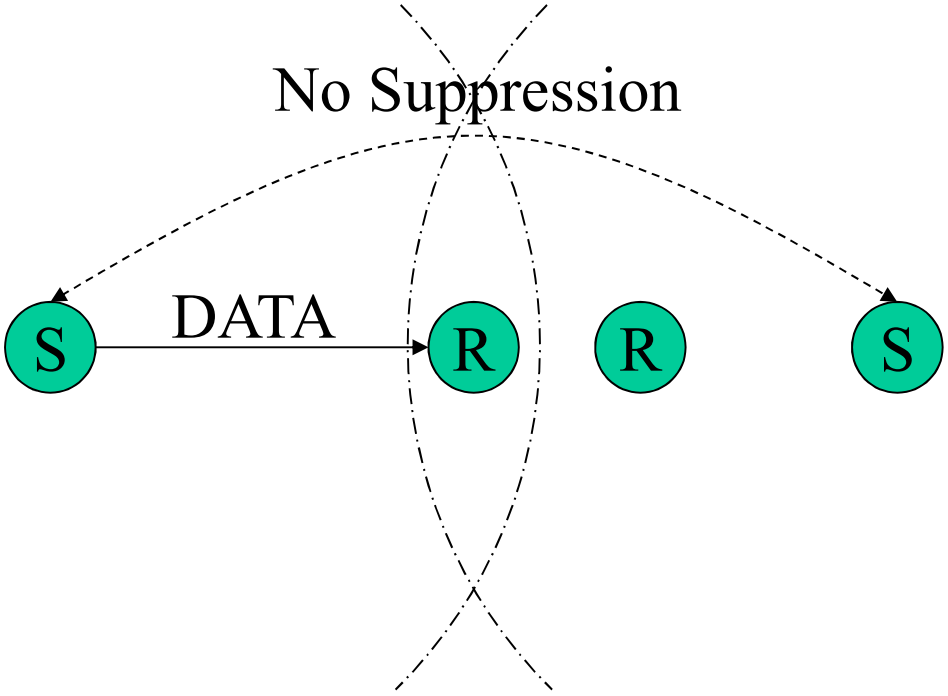
CSMA/CA and ACK Suppression by Carrier Sense



CSMA/CA and ACK Suppression by Carrier Sense



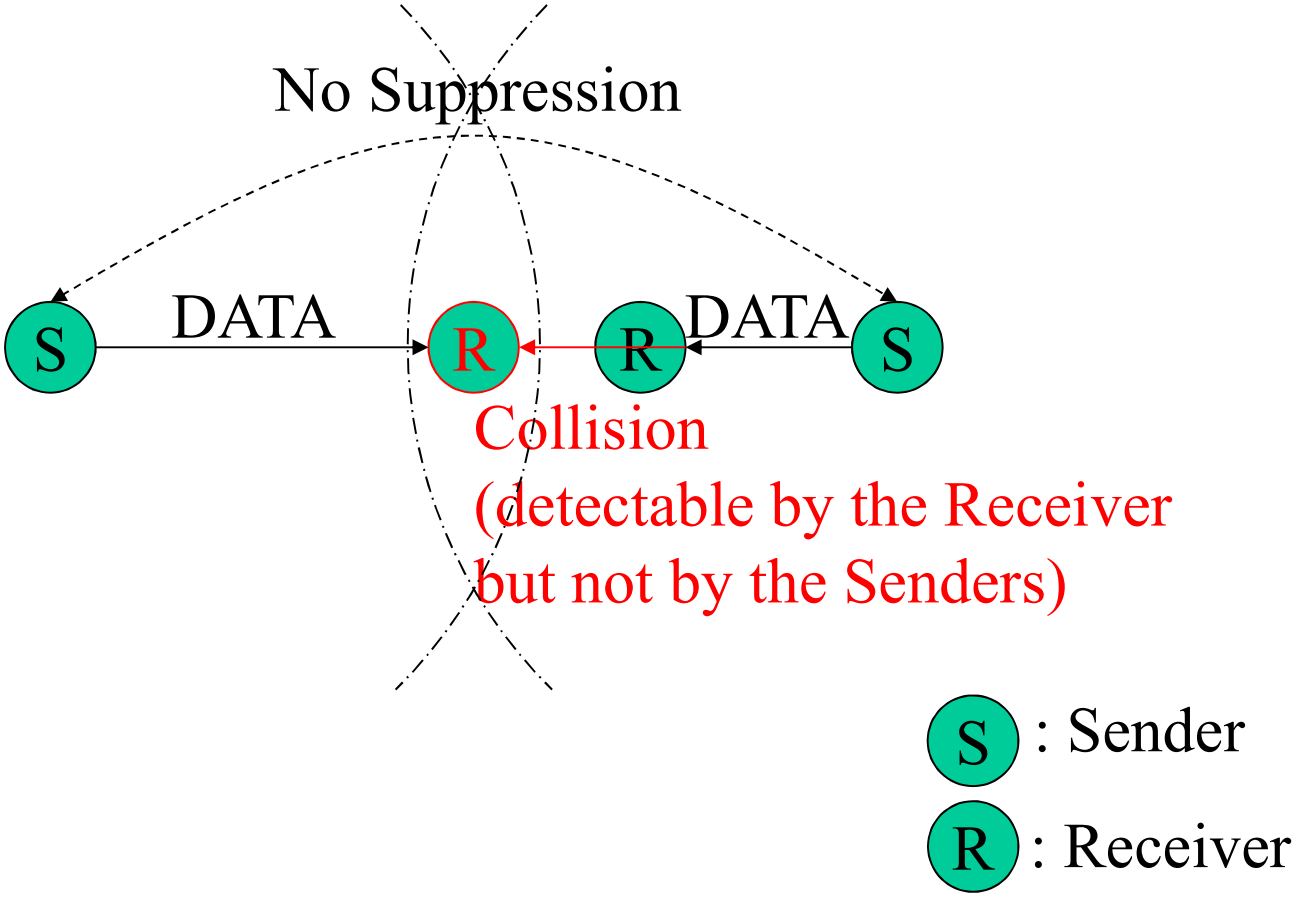
CSMA/CA and ACK
Collision by Hidden Terminal



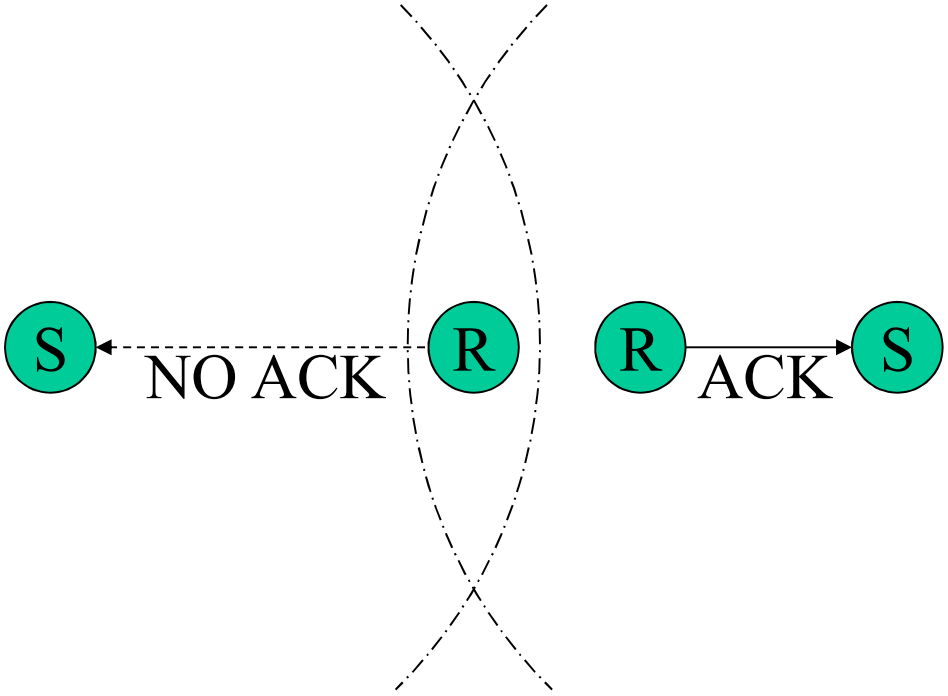
S : Sender

R : Receiver

CSMA/CA and ACK
Collision by Hidden Terminal



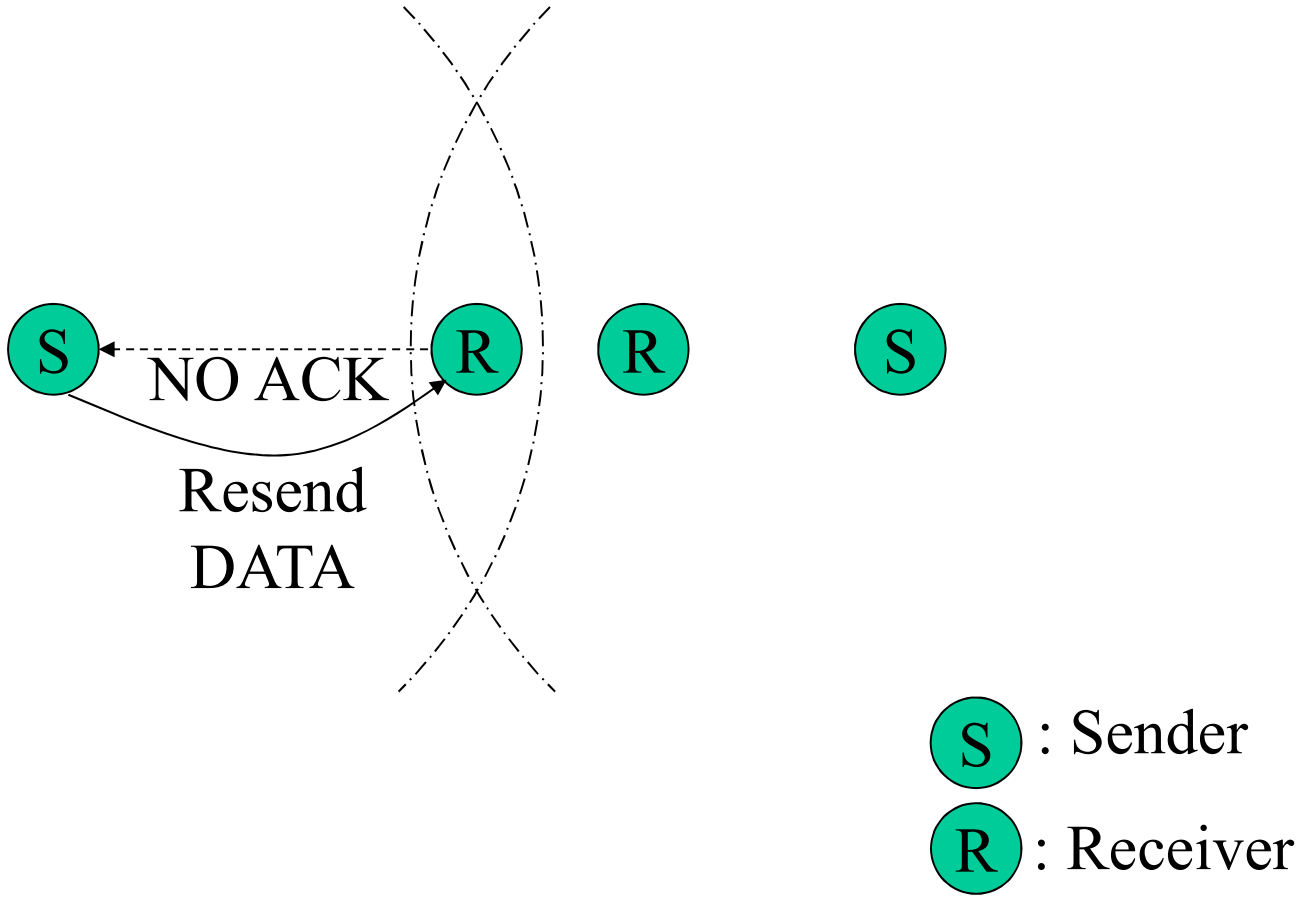
CSMA/CA and ACK
Collision by Hidden Terminal



S : Sender

R : Receiver

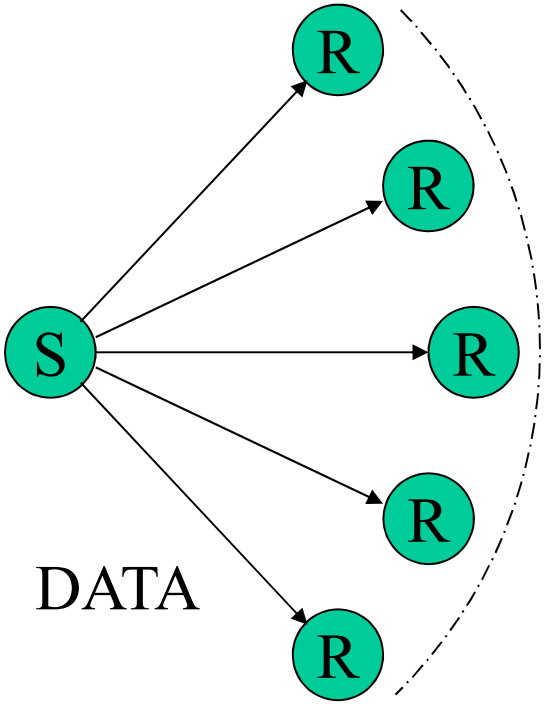
CSMA/CA and ACK
Collision by Hidden Terminal



Wireless LAN of IEEE 802.11

- Relies on CSMA/CA
  - because of undetectable collisions
    - ACKs are the MUST for reliable communication
      - or packets are lost upon collisions
    - Unicast packets are ACKed and delivered reliably
  - Broadcast/multicast packets ***can not*** be ACKed
    - Broadcast/multicast packets are delivered ***unreliably***
      - ***The major difference to Ethernet***

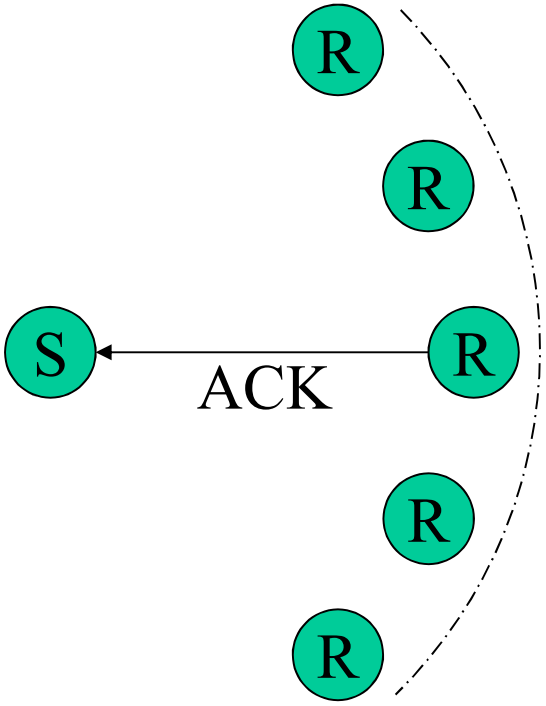
CSMA/CA and ACK
Multicast and Lack of ACK



S : Sender

R : Receiver

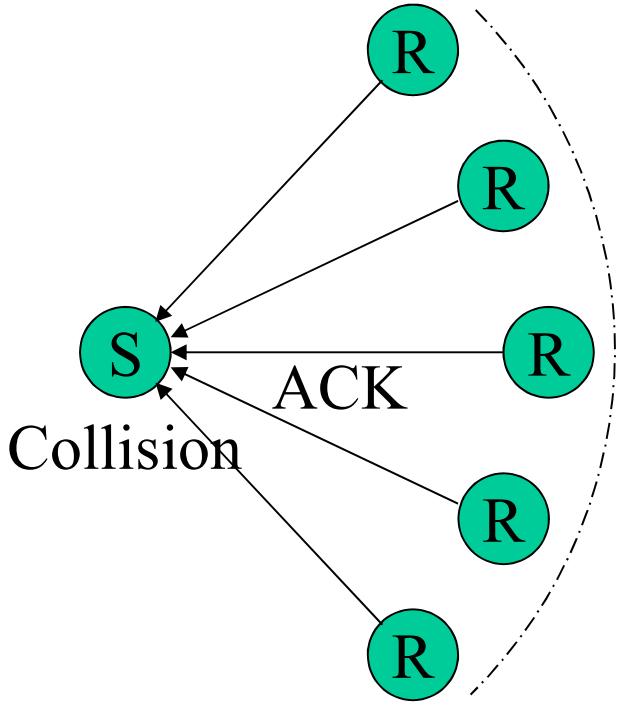
CSMA/CA and ACK
Multicast and Lack of ACK



S : Sender

R : Receiver

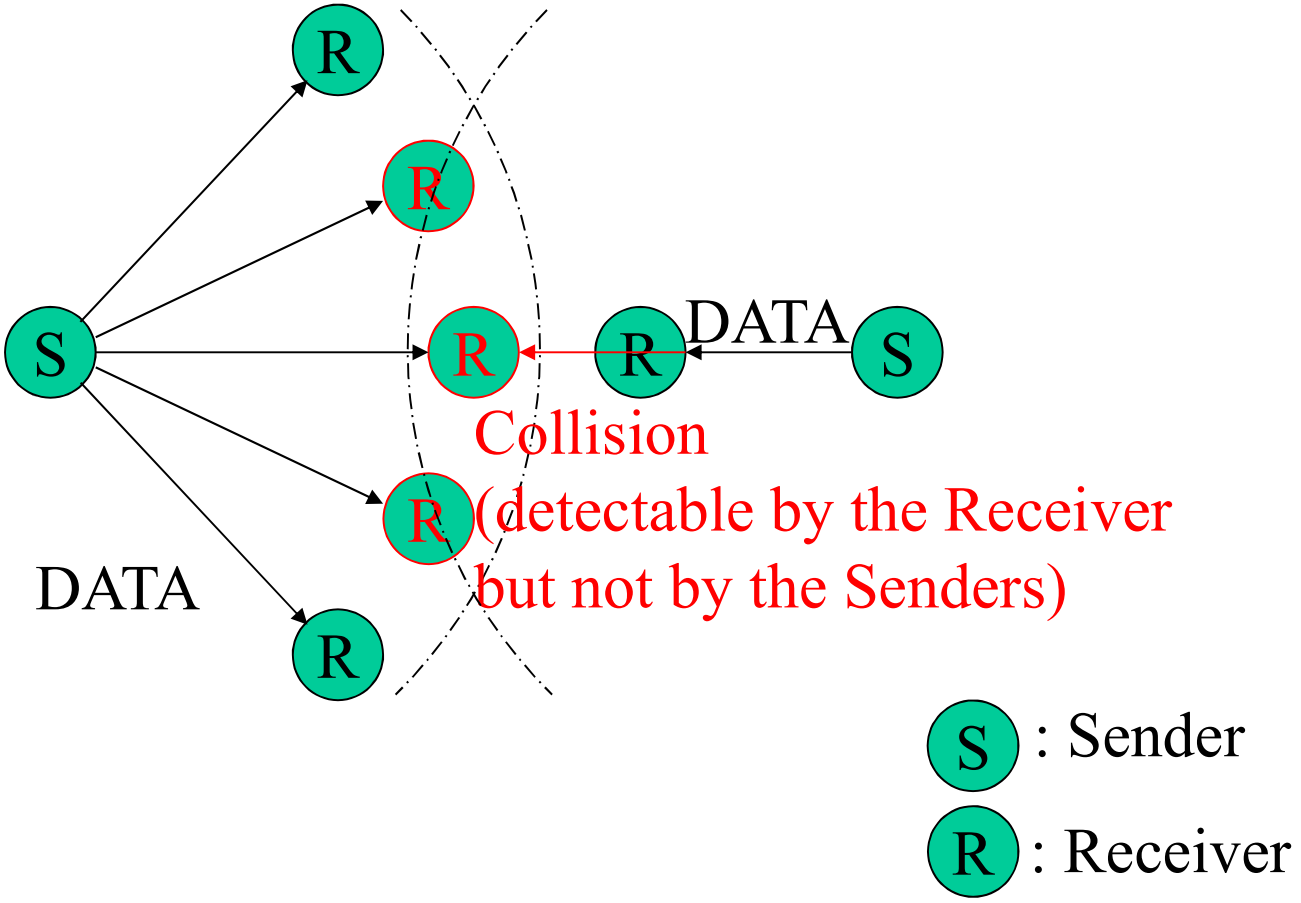
CSMA/CA and ACK
Multicast and Lack of ACK



S : Sender

R : Receiver

CSMA/CA and ACK
Multicast and Unreliability



Wireless LAN of IEEE 802.11

- Relies on CSMA/CA
  - because of undetectable collisions
    - ACKs are the MUST for reliable communication
      - or packets are lost upon collisions
    - Unicast packets are ACKed and delivered reliably
  - Broadcast/multicast packets ***can not*** be ACKed
    - Broadcast/multicast packets are delivered ***unreliably***
      - ***The major difference to Ethernet***
  - Reliable broadcast is by ***frequent*** beacons

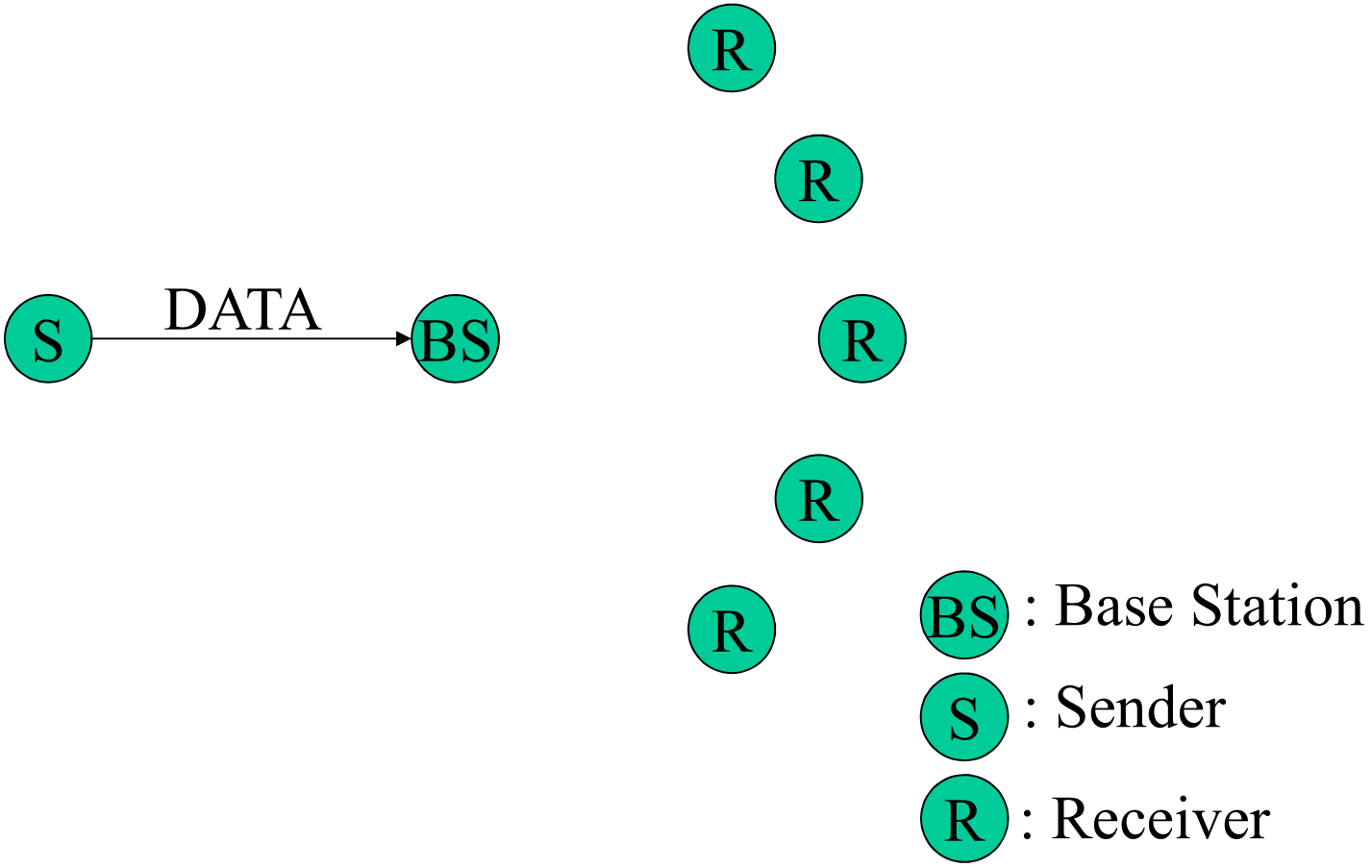
Reliability by Frequent Beacons

- If broadcast is received 20% of the time
  - repeated beacons will finally be received
    - with 10 repetitions, 90% of the time
    - with 20 repetitions, 99% of the time
- If broadcast is received 10% of the time
  - repeated beacons will finally be received
    - with 10 repetitions, 65% of the time
    - with 20 repetitions, 88% of the time
    - with 40 repetitions, 99.5% of the time

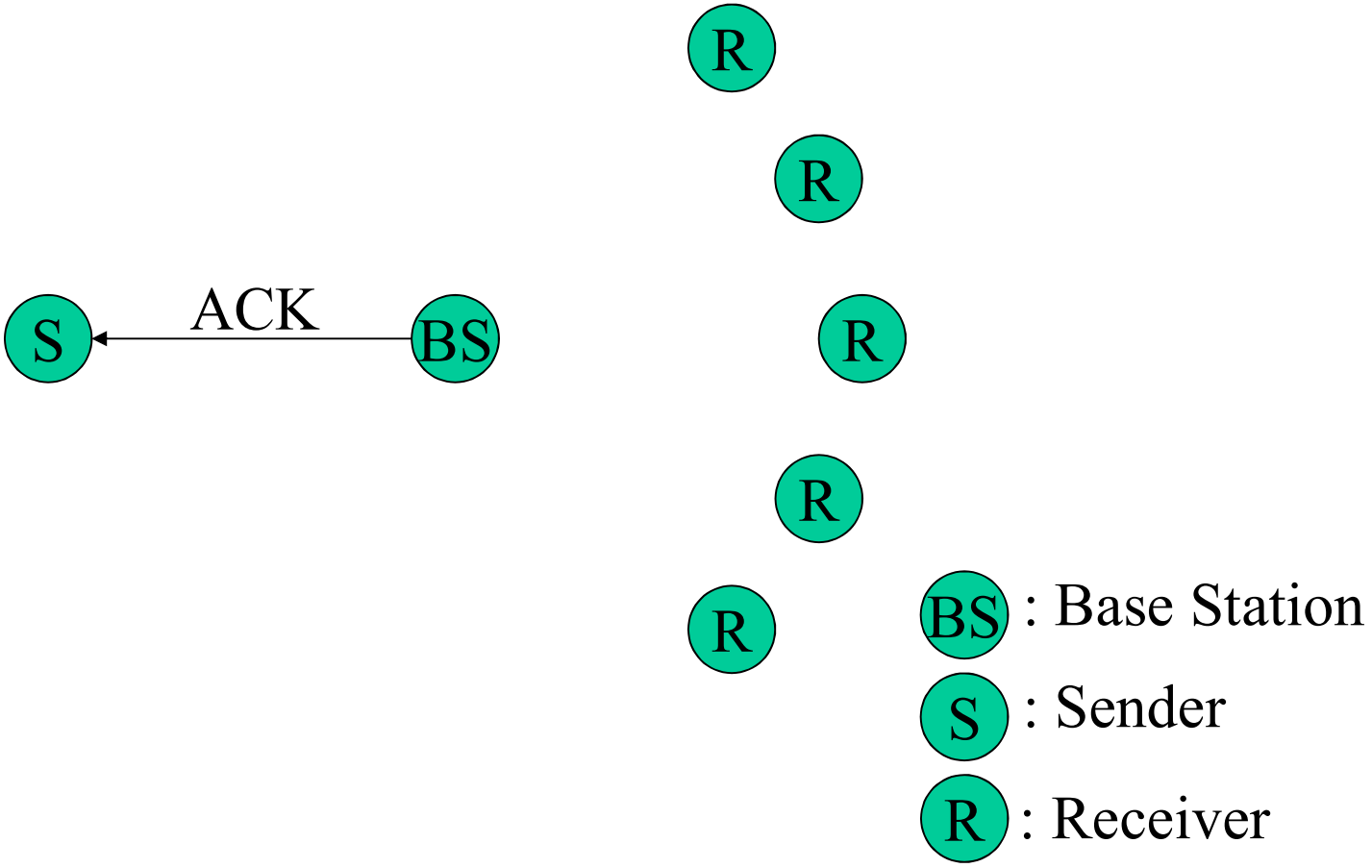
Broadcast over Wireless LAN of Infrastructure IEEE 802.11

- Stations (STAs) send broadcast packets to the base station (BS) through link unicast
  - delivery is ACKed and reliable
    - Broadcast from STAs is received by BS reliably
  - BS, then, broadcast the packet to all the STAs
    - Broadcast from the STAs to non-BS STAs are unreliable

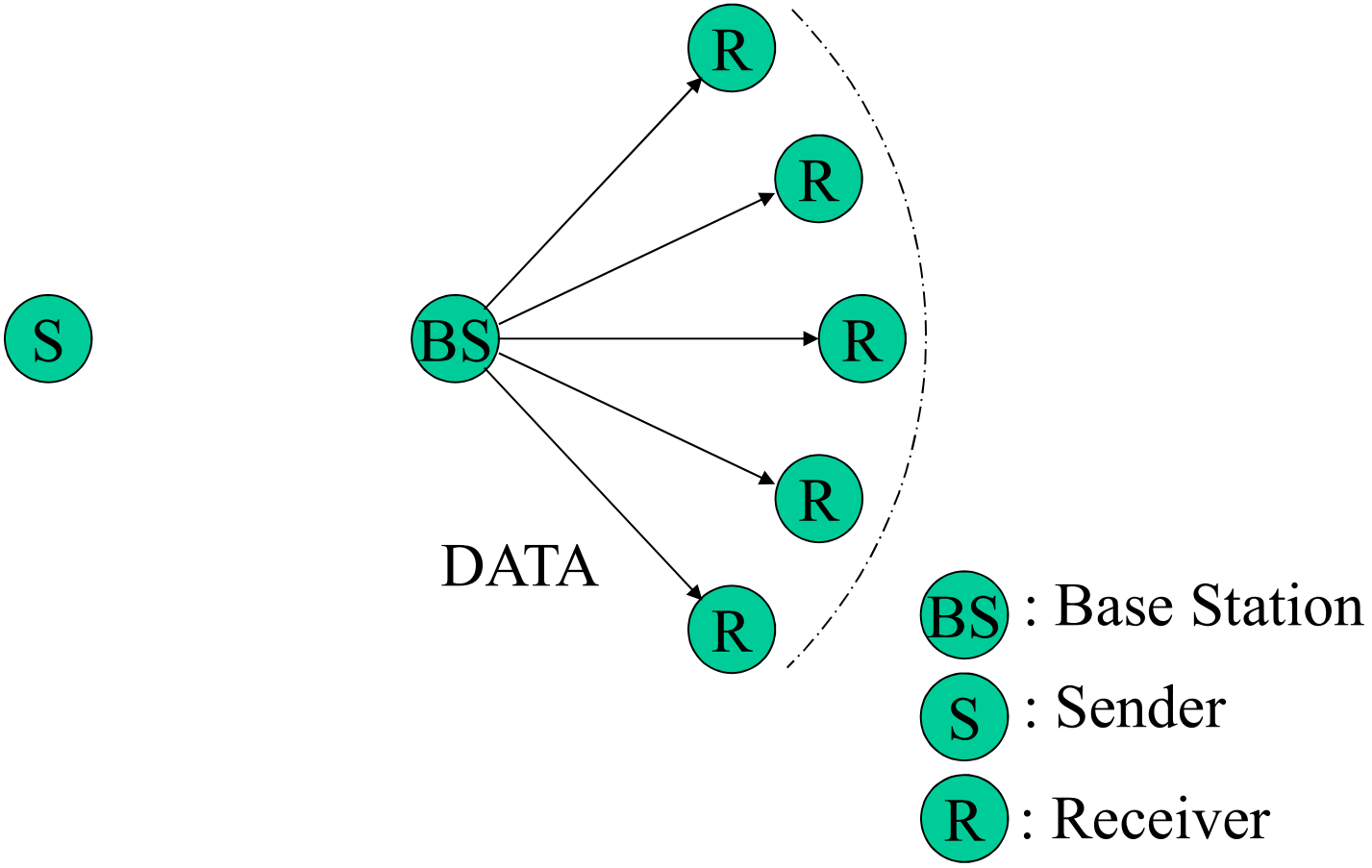
Broadcast/Multicast over Infrastructure WLAN



Broadcast/Multicast over Infrastructure WLAN



Broadcast/Multicast over Infrastructure WLAN



ARP
The Way for IP over Ethernet

- IP uber alles! (IP over everything!)
  - IP ***MUST*** work over any link layers
    - Various adaptation mechanisms take care of matching between L3 and L2s
      - The adaptation mechanisms take care of differences between various L2s
    - ARP (of IPv4) is the adaptation mechanisms between IP and Ethernet

Neighbor Discovery (ND)
The Major Design Flaw of Ipv6

- ~~IP uber alles! (IP over everything!)~~
- ND uber alles!? (ND over everything)
  - ND ***MUST*** work over any link layers!?
    - A single adaptation mechanism ***CAN NOT*** take care of matching between L3 and various L2s
      - The single adaptation mechanism ***CAN NOT*** take care of differences between various L2s
  - ND was designed for Ethernet, PPP and ATM
    - but not for Wireless LAN nor other L2s
      - ND ***MAY NOT*** be able to take care of Wireless LAN

Wrong Assumptions of ND on L2

- The world will be ATM centric
  - IP over a large L2 cloud of worldwide ATM
    - L1/L2 broadcast is inhibited
      - timeout period of L2 multicast (P2MP) is long
- Terminals are mostly immobile
  - "Routers generate Router Advertisements frequently enough that hosts will learn of their presence ***within a few minutes***" (RFC2461)
- L2 broadcast/multicast is reliable

The Reality of L2s under IP

- The world is IP centric
  - ATM has gone
- L2 is small
  - The CATENET model, of course
- Terminals are highly mobile
  - can't wait a few minutes for network reconf
- L2 broadcast/multicast is ***UNRELIABLE*** over (congested) WLAN

How ND was expected to Work over WLAN

- NS (Node Solicitation) is multicast
- RS (Router Solicitation) is multicast
  - received unreliably (except by BS)
  - Then, RA (Router Advertisement) is unicast/multicast
- Unsolicited RA is multicast

Other Protocols Affected

- Protocols using broadcast/multicast suffer
  - DHCP, ARP, Routing Protocols, ...,
- However, if BS is the only router
  - DCHP discover to BS is reliable
  - ARP to BS is reliable
  - ARP from BS is unreliable
    - not common in ad hoc environment
  - Routing protocols are not necessary

Reaction from IETF

- The problem is recognized
- Treat WLAN as NBMA?
  - However, RFC2461 says “The details of how one uses ND on NBMA links is an area for further study.”

Conclusions

- Wireless LAN and Ethernet are different
- “ND over everything” is a bad idea
  - proven by the most popular (next to Ethernet) L2 technology of wireless LAN
- Further study is necessary
  - to make IPv6 deployable
- IP uber alles!
  - not necessarily IPv6

How Path MTU Discovery not Work

Masataka Ohta
Tokyo Institute of Technology
mohta@necom830.hpcl.titech.ac.jp

Abstract

- Multicast path MTU discovery (PMTUD) is a new feature of IPv6. However, ICMP implosion with multicast PMTUD can be serious when most MTU bottlenecks are located near individual receivers. ICMP Packet Too Big, at least those generated against multicast packets, will be filtered, which is a standard violation, which means there is no reason not to filter unicast ones. Thus, unicast PMTUD is not expected to work. We should not send packet >1280B, except for IP over IP tunnels.

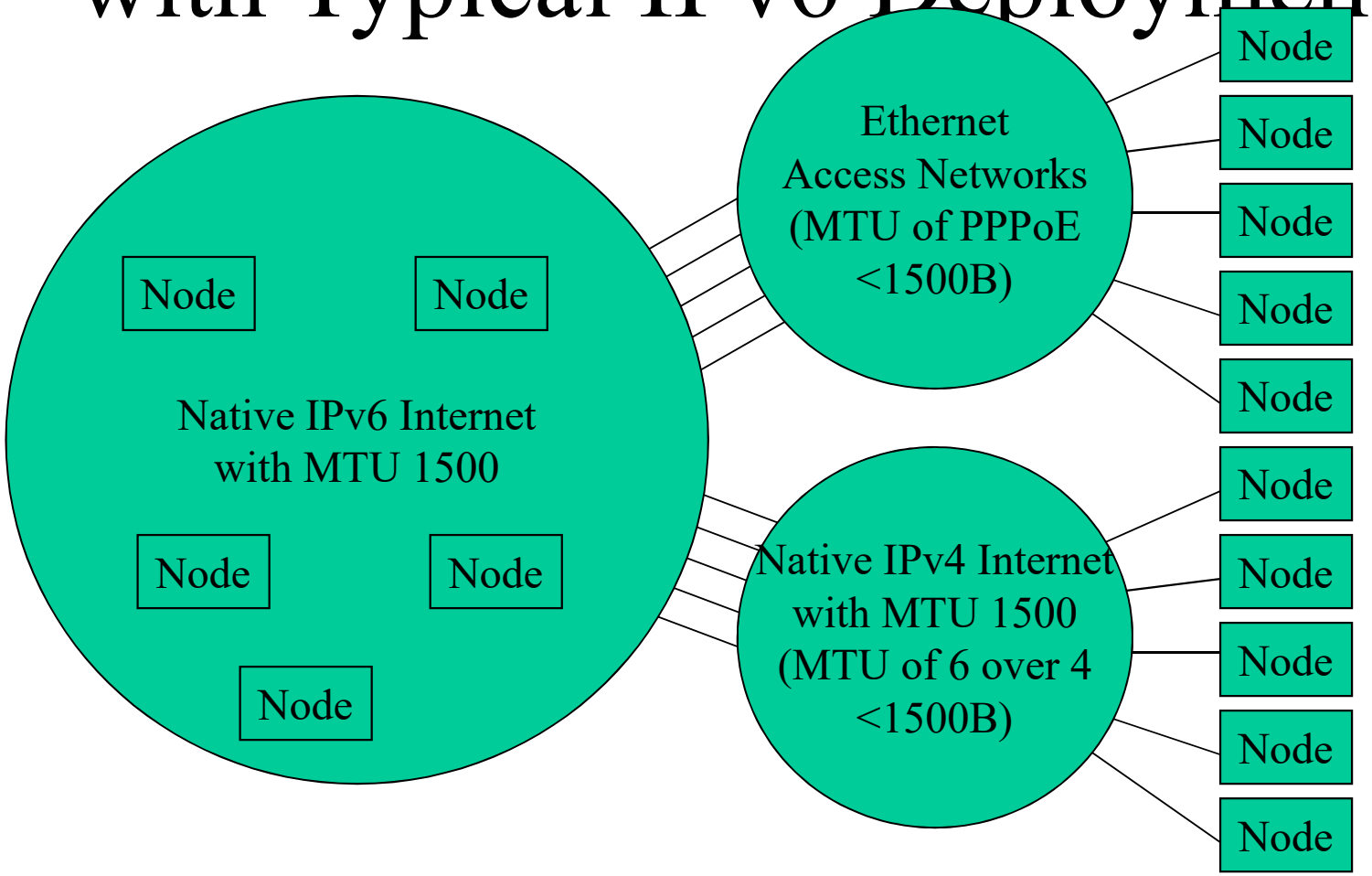
PATH MTU Discovery

- Measure Path MTU by ICMP Packet Too Big
  - Path MTU is set to the value contained in the ICMP packet
  - does not work if ICMP Packet Too Big is filtered or not generated
- Periodically send larger packet to detect MTU increase by path change
- “SHOULD be supported” (node requirement)

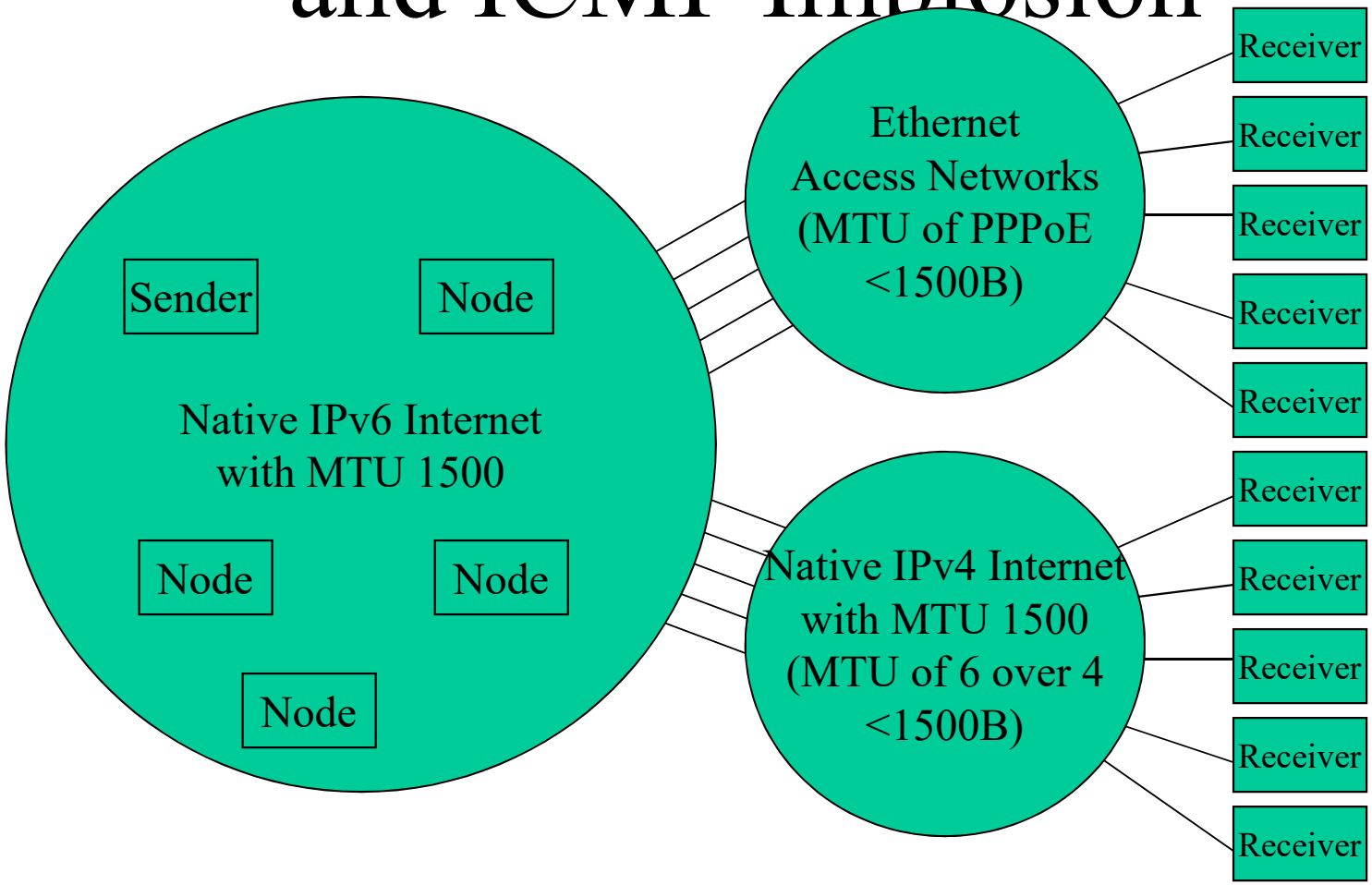
RFC1981 (Path MTU Discovery for IP version 6)

- The Draft Standard Specifies:
  - Path MTU Discovery supports multicast as well as unicast destinations. In the case of a multicast destination, copies of a packet may traverse many different paths to many different nodes. Each path may have a different PMTU, and a single multicast packet may result in multiple Packet Too Big messages, each reporting a different next-hop MTU. The minimum PMTU value across the set of paths in use determines the size of subsequent packets sent to the multicast destination.
  - In the case of a multicast destination address, copies of a packet may traverse many different paths to reach many different nodes. The local representation of the "path" to a multicast destination must in fact represent a potentially large set of paths.
- How large is “a potentially large set of

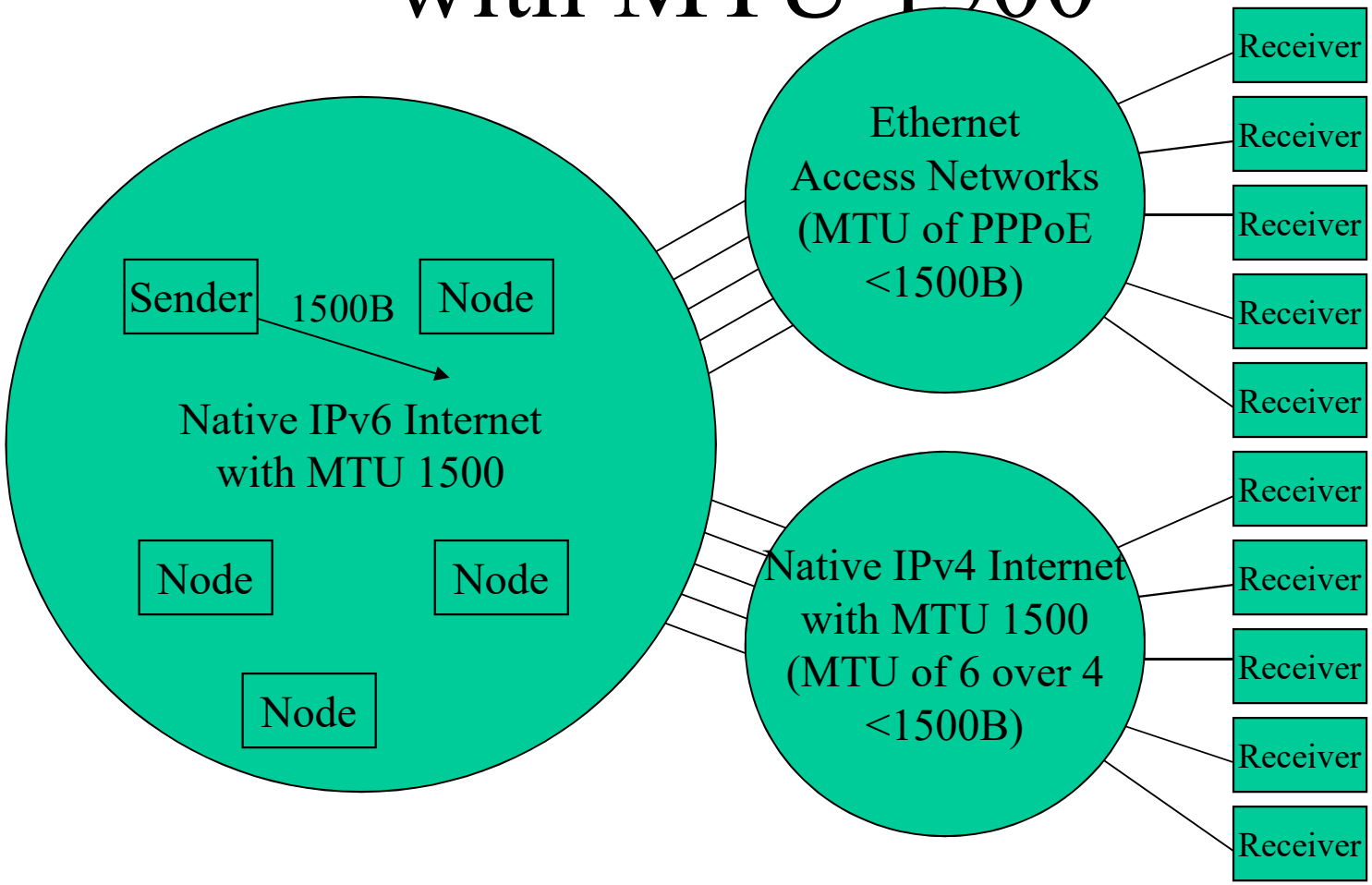
Tunnels at the Last Hop with Typical IPv6 Deployment



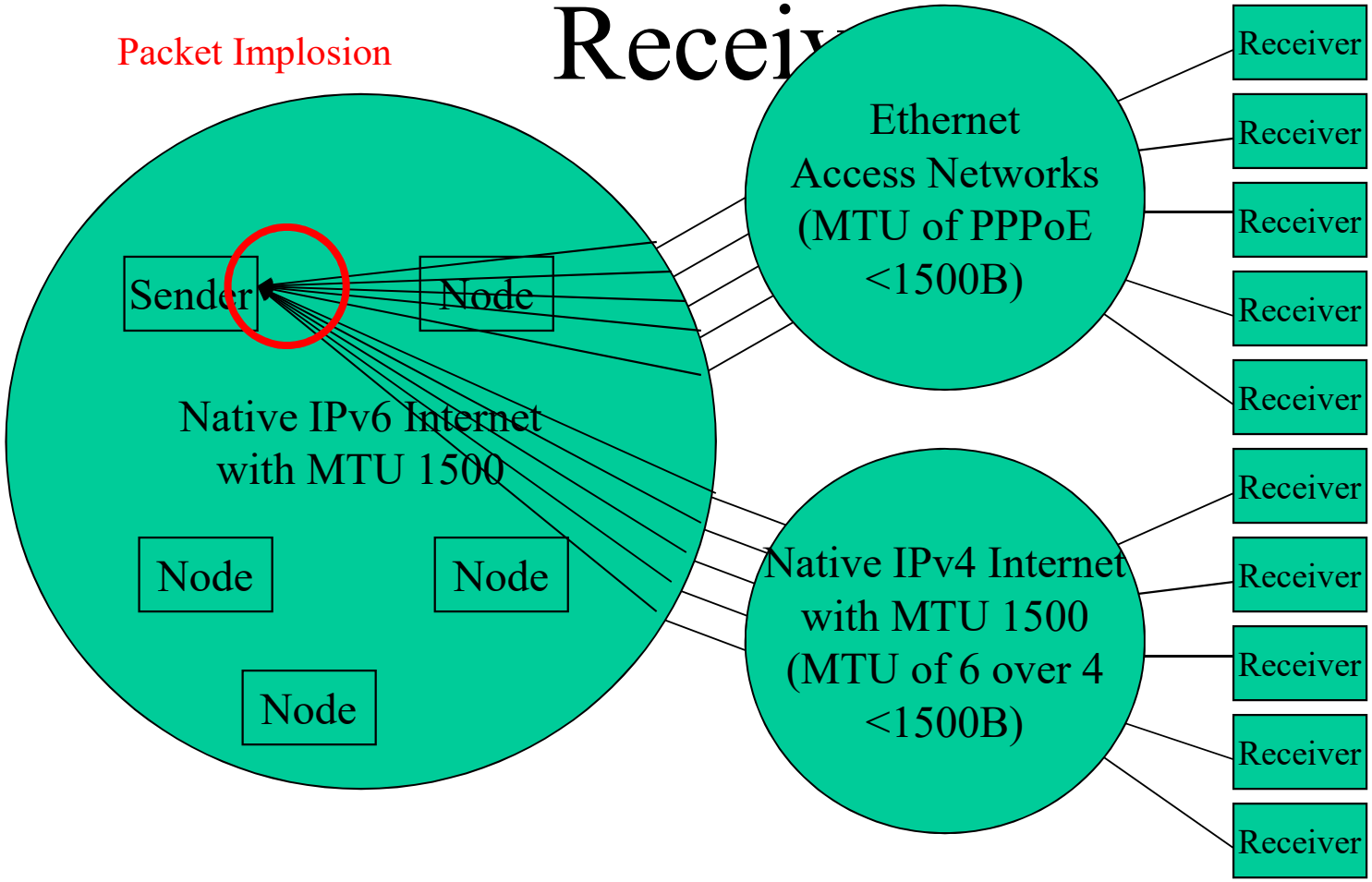
Multicast Path MTU Discovery and ICMP Implosion



Sender Periodically Send Packets with MTU 1500



ICMP Packet Too Big Messages are Generated Near Each Receiv



DOS

- Some multicast routing protocol allows for source address spoofing
  - ICMP may be used for DOS amplifier
  - even if non-link-local multicast is not enabled around a victim

Not a Problem?

- Because almost all ISPs do not enable multicast routing protocol
- ISPs do not allow ordinary users send multicast packets
  - still a problem, because rational ISPs want to avoid to rely on rational operations of other ISPs
  - instead, the multicast PMTUD problem is yet another reason for ISPs to disable multicast
    - multicast PMTUD, to promote multicast and

RFC2463 (ICMPv6) Requires

- A Packet Too Big MUST be sent by a router in response to a packet that it cannot forward because the packet is larger than the MTU of the outgoing link. The information in this message is used as part of the Path MTU Discovery process [PMTU].
- Sending a Packet Too Big Message makes an exception to one of the rules of when to send an ICMPv6 error message, in that unlike other messages, it is sent in response to a packet received with an IPv6 multicast destination address, or a link-layer multicast or link-layer broadcast address.
  - Parameter Problem Messages also make an exception

To Prevent ICMP Implosions

- Violate RFC2463 to
  - stop generating ICMP packet too big and parameter problem for multicast packet
  - filter ICMP packet too big and parameter problem for multicast packet
- Or, as it is already a violation, simply
  - stop generating any ICMP
  - filter all the ICMP
  - “it’s against an RFC” is not a valid criticism

Fundamental Solution

- Update RFC2463 to prohibit generation of ICMP against multicast packets
- Write an BCP to Force ISPs not Filter ICMP
- Should take another decade or two
  - unrealistic

Without PMTUD...

- According to RFC2460:
  - It is strongly recommended that IPv6 nodes implement Path MTU Discovery [RFC-1981], in order to discover and take advantage of path MTUs greater than 1280 octets. However, a minimal IPv6 implementation (e.g., in a boot ROM) may simply restrict itself to sending packets no larger than 1280 octets, and omit implementation of Path MTU Discovery.
    - Packet larger than 1280B can not be sent
- IP over IP tunnels (e.g. RFC2473 for MIPv6) needs tunnel MTU 1280B, violating RFC2460
  - or, all the 1280B packets are fragmented,

Conclusion

- Multicast PMTUD is broken
  - to cause ICMP implosion
- ISPs should filter ICMP Packet Too Big
  - at least against multicast packets but maybe all
- We can’t expect unicast PMTUD work
- We shouldn’t send packets > 1280B
  - except for tunnels

新世代ネットワークアーキテクチャ実現に向けて
—AKARIプロジェクトからの報告—

新世代ネットワークワークショップ

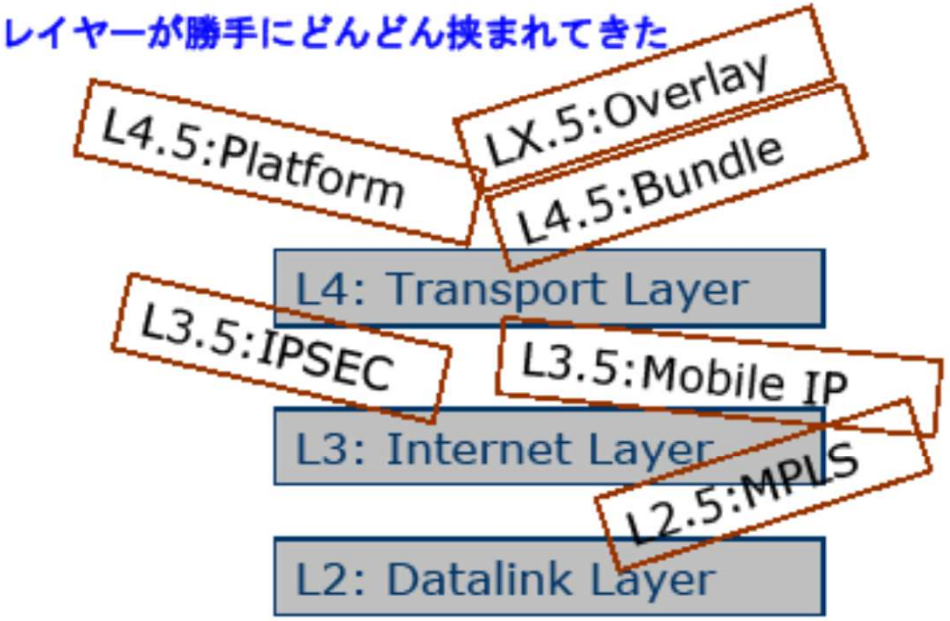
2007年6月11日

平原 正樹
ネットワークアーキテクチャグループ
新世代ネットワーク研究センター
NICT

インターネット – あまりにも複雑/矛盾 → 破綻 第２章１節

新しい機能を積み上げることができない。未来の社会を支えるサービスを提供できない。

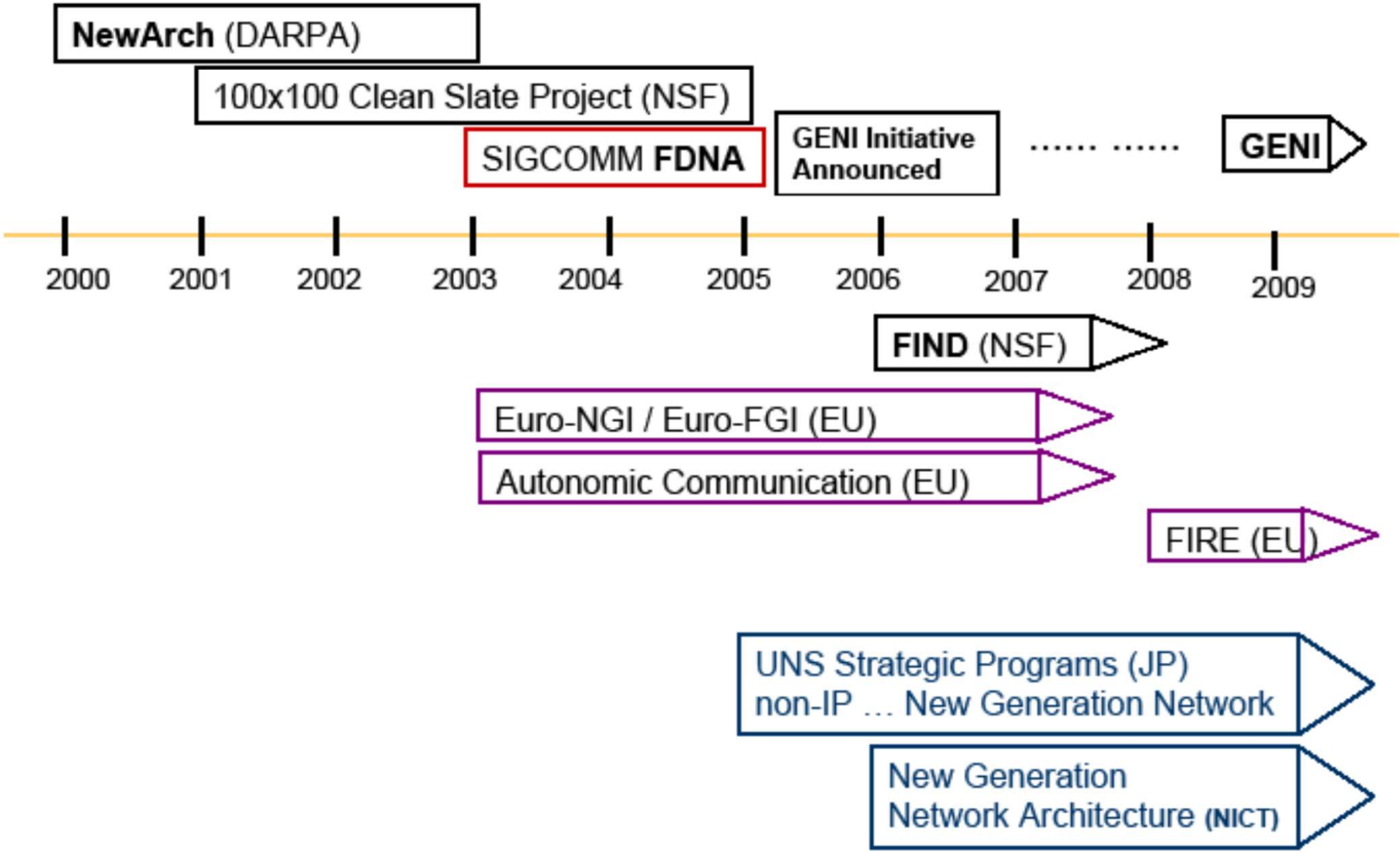
・命を預けられるか?(遠隔医療、交通、緊急通報)
・生活を預けられるか?(防犯、契約行為、金融)
・生活を豊かにできるか？(センサー、RFID)
・安心した生活が送れるか?(対スパム、耐攻撃)
・あと何十年使えるの？(インフラ、持続社会)
・未来の変革を受け入れる余裕は?(未知の要求)



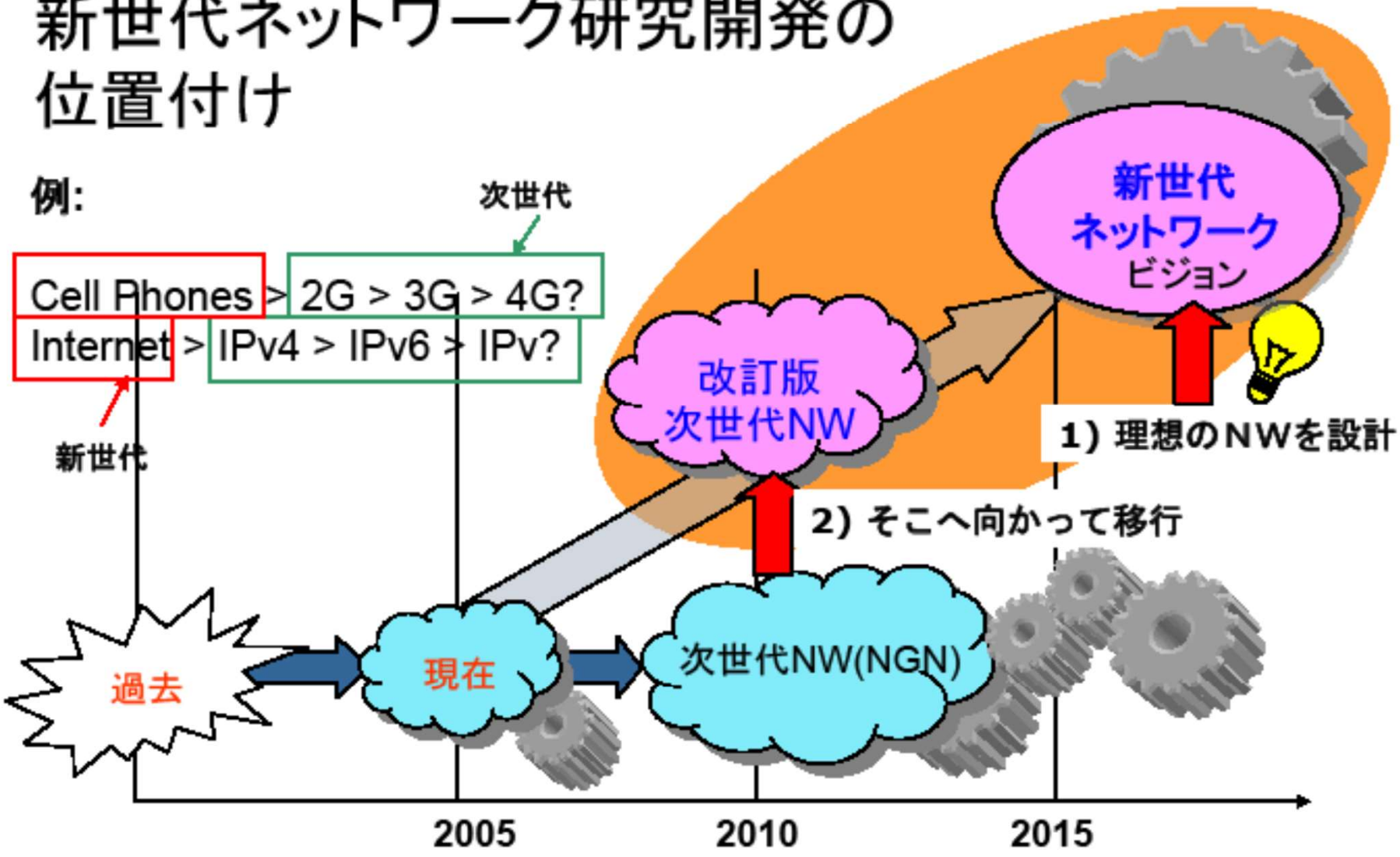
個々はＯＫでも全体では×

一から作り直すべき時期が迫っている！

ネットワークアーキテクチャ“白紙から作り直し”機運



新世代ネットワーク研究開発の位置付け



AKARIプロジェクト

- a small light in the dark pointing to the future -

目標: 2015年の新世代ネットワーク設計図

・現在のしがらみに捕われない。・白紙から理想を追い求める。・その後で現在からの移行を考える。

グループリーダー:平原 正樹
原井 洋明(光交換)、徐 蘇鋼(光パス)、宮澤 高也
盛岡 敏夫(光伝送)、大槻 英樹(ネットワーク制御)、Jumpot Phuritatkul
井上 真杉(アクセス)、中内 清秀(オーバーレイ)、Ved Kafle (アドレッシング)

客員研究員

大阪大学 村田教授 (ネットワーク科学)
慶応大学 寺岡教授 (モビリティ)
東京大学 森川教授 (ユビキタス)
東京工業大学 太田講師 (パケット交換)

アドバイザ:
青山(プログラムディレクタ)
久保田(センター長)

ミーティング 2回/月, 合宿 2回/年

AKARI アーキテクチャの目的

持続可能なネットワークアーキテクチャ

量から質へ (Capacity for Quality) **簡約化 (KISS)**

ネット空間の現実化 (Realizable Network Space)

人類の可能性を伸ばす (Future Diverse Society)

新世代ネットワークアーキテクチャの原理原則

1. KISS原則 (Keep It Simple, Stupid)
   - 結晶合成 (選択・統合・単純化)
   - 共通レイヤ (レイヤ縮退)
   - End-to-End (Original Internet)

2. 現実結合原則
   - 物理・論理アドレス分離
   - 双方向認証
   - 追跡可能

3. 持続的な進化原則
   - 自己創発(エマージェント)
   - 自律分散制御
   - スケーラブル
   - 社会選択

「パケット交換」って、具体的に何？

- PDMA（第二回）
- 光パケット（最終回）
- IP－－（今回）

IP－－

- IPv6の本来あるべき姿
  - アドレス空間の拡張
  - 経路表の徹底的階層化
  - 単純化

IP－－の特徴

- IPオプション無し
  - ヘッダは、15個までのソースロケータを含む
- 最小MTU9KB、PMTUD無し
- ブロードキャストのサポート
- マルチキャスト、IPSECはオプション
- IDとロケータの分離
  - ロケータは12ビット単位で5階層
    - 全ホストは5階層分の経路表を持つ

IP－－のパケットフォーマット

ペイロード長	プロトコル	HTL
ソースLocator List長	未使用	
ソースLocator		
ソースID		
ディスティネーションLocator		
ディスティネーションID		
ソースLocator List		
ペイロード		