

## 研究の背景

- NICT では新世代ネットワーク設計のための AKARI プロジェクトを推進している
  - 一から設計し直すという方針 (a clean-slate approach)
- Clean-slate なアドレスリナンバリングとネットワーク自動構築について考察
  - アドレス空間としては 128bit 長程度を考慮
  - ただし本スライドでは読み易さを考慮して IPv4 の表記を流用

## 階層的なロケータ自動番号割当プロトコル HANA の設計と実装

藤川 賢治 (FUJIKAWA Kenji) <hudikaha@nict.go.jp>  
(独)情報通信研究機構 (NICT)

太田 昌孝 (OHTA Masataka) <mohta@necom830.hpcl.titech.ac.jp>  
東京工業大学 (TITECH)

## 本研究の目的

1. End-to-end マルチホームによるインターネットバックボーンの経路表の削減
  - ISPに縛られないようにする為アドレスPrefixは上流ISPから自動割当されるべき
2. ルータの設定の手間を最小化
  - Suffixは自動割付けされるべき
3. Prefix設定とSuffix設定を統合
  - DNS へのアドレス登録も含む

## 現在のインターネットの経路表とマルチホーム

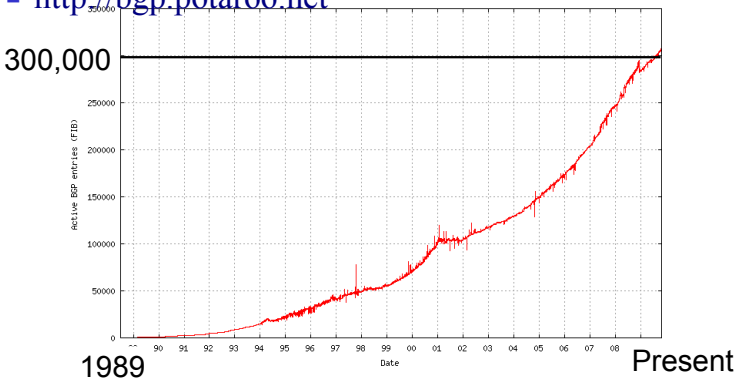
- インターネットバックボーンの経路表は増大を続けている
- 経路収束時間の劣化は深刻な問題
- マルチホームが問題を悪化させている
  - サイトは冗長性確保の為マルチホームを要求

→End-to-end マルチホームが本問題を解決

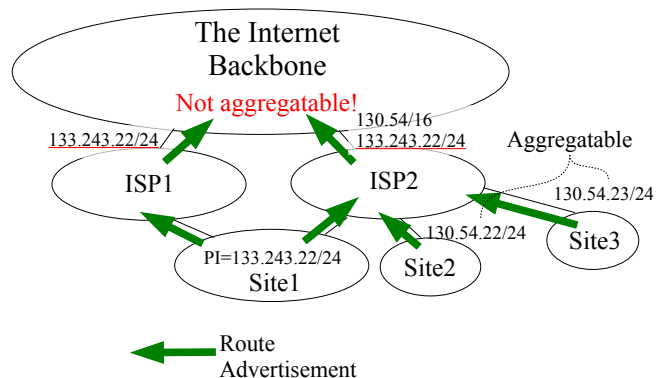
## →Hierarchical Automatic Number Allocation (HANA) プロトコルを設計

## Active BGP Entries from 1989 to Present

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

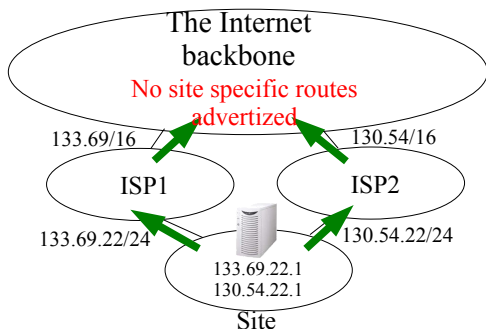


## PI (Provider Independent) アドレスを用いた BGP によるマルチホーム



Prefixの50%以上が/24!

## PA (Provider Aggregatable) アドレスによる



PIアドレスを認めた IPv6 では問題の解決にならない

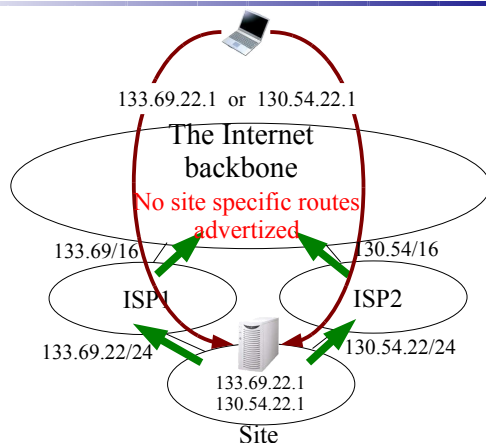
- /24 が大域経路表の 50%を占める
- ということは PIアドレスを認めてい IPv6 では良くて経路表は半分にしかならないと予想される

## End-to-end マルチホーム

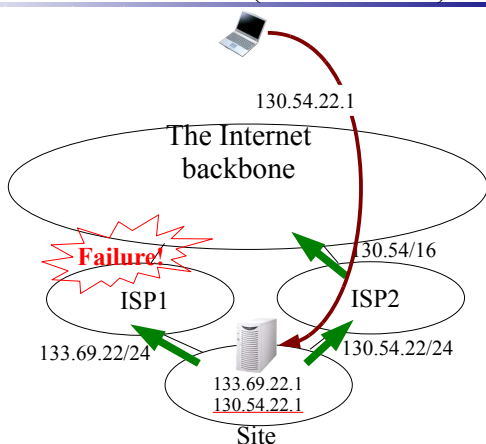
- M. Ohta and M. Sola, "The Architecture of End to End Multihoming," IETF Internet-Drafts (already expired), draft-ohta-e2e-multihoming-00.txt, 28 Apr 2000.

> This memo describes the architecture of end to end multihoming. End to end multihoming does not burden routing system for multihoming. That is, even extensive use of end to end multihoming does not increase the number of entries in a global routing table. Traditionally with IPv4, multihoming capability is offered by an intelligent routing system, which, as is always the case with violating the end to end principle, lacks scalability on a global routing table size and robustness against link failures. On the other hand, **with end to end multihoming, multihoming is supported by transport (TCP) or application layer (UDP etc.) of end systems** and does not introduce any problem in the network and works as long as there is some connectivity between the end systems. Because end to end multihoming is performed in end systems, the architecture needs no routing protocol changes. Instead, APIs and applications must be modified to some extent.

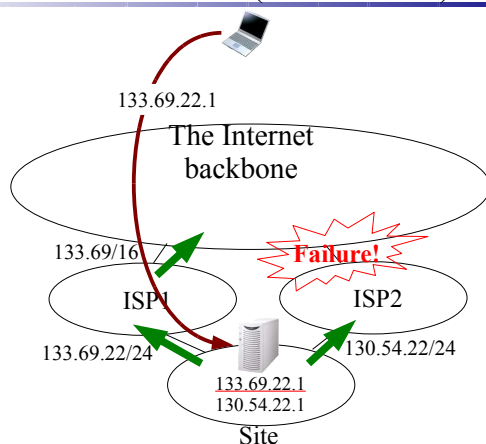
## End-to-end マルチホームにおける相手側エンドによるアドレス選択



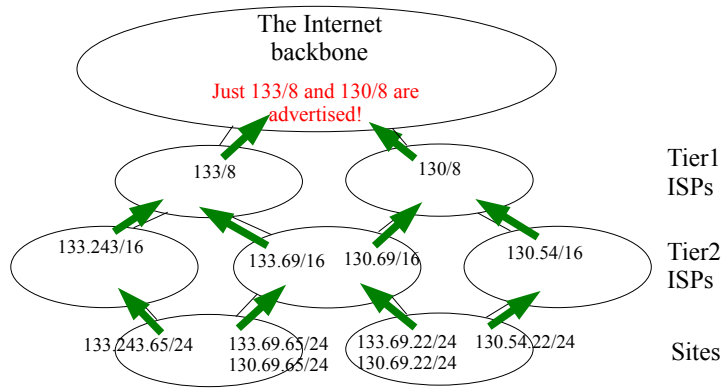
## End-to-end マルチホームにおける相手側エンドによるアドレス選択 (ISP1 に障害)



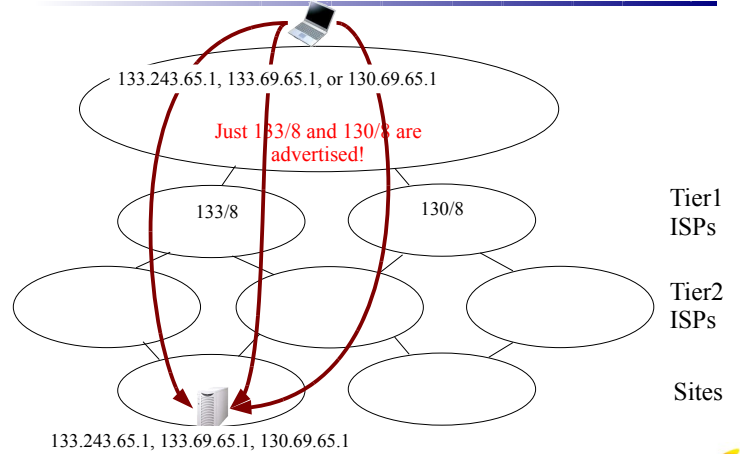
## End-to-end マルチホームにおける相手側エンドによるアドレス選択 (ISP2 に障害)



## マルチホームにおける階層的なアドレス割当てと広告



## End-to-end マルチホームにおける相手側エンドによるアドレス選択



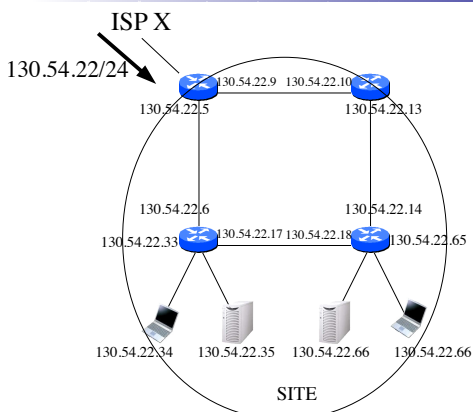
## End-to-end マルチホーム実現に向けた課題

1. 末端ホストが複数アドレスを用いた接続を確立する手法の提供
  - 複数アドレスを接続確立時に試していくプログラミング作法の普及
    - SMTPはDNS MXレコードにより解決
    - それ以外でもIPv6に対応しているのなら同様の動作
  - 複数アドレスを利用するTCPの開発と普及
    - SCTPの普及?
2. ソースアドレス選択
3. リナンバリングの提供
4. DNS自動登録

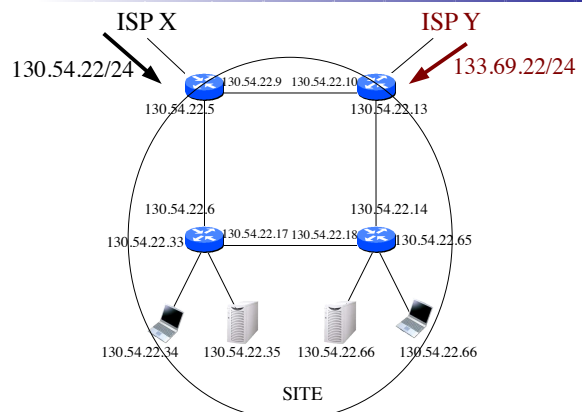
## 研究の目的 1: マルチホームの為の新しいリナンバリングアーキテクチャを設計

- 基本方針
  - アプリケーションは DNS名を用いて相手を指定
    - アドレスを直接使わない
  - DNSデータベース更新機構も考慮の必要あり
  - 階層化とマルチホームを考慮した自動リナンバリングプロトコルの開発
- 本研究では自動リナンバリングに関して議論

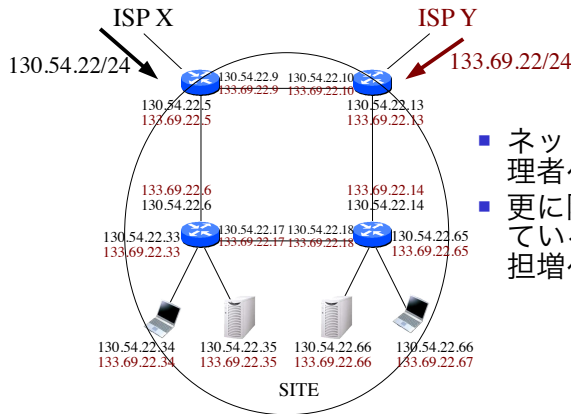
## Prefix 配布: 上流が一つのサイト



## Prefix 配布: 別のISPにも接続



## Prefix 配布: 新しいPrefixの配布



- ネットワーク管理者への負担増
- 更に階層化されている更なる負担増へ!

## 研究の目的 2: Suffix の自動割付

- ルータの設定の手間を最小化
- L2スイッチをL3ルータで置換
  - QoS保証光パスを提供するL3ルータ別途開発中
  - 光パスを提供する L3 ルータが末端ホストの直前に置ければ、end-to-end光パスが実現可能
  - この為にはL3ルータはL2スイッチ並に自動設定できる必要あり。末端利用者は設定の必要な箱は欲しくない

## 既存研究: IPv6 Router Renumbering (RR) プロトコル

- IPv6 RR プロトコル (RFC 2894) が開発されたが広く使われているわけではない
- 以下の機能が考慮されていない
  - DNS データベースの変更
  - アプリケーションに埋込まれたアドレスの変更
  - Suffix の取扱い

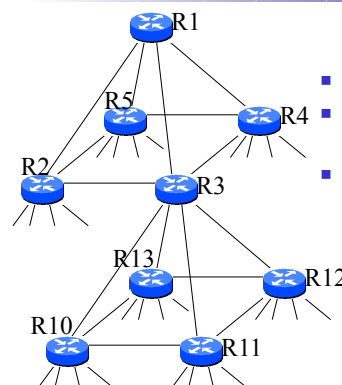
## Hierarchical Automatic Number Allocation (HANA) プロトコル

- 階層化リナンバリング機能の提供
  - Prefixは階層的に上流から通知される
- ルータ自動設定機能の提供
  - SuffixはDHCPのような方法で自動設定される
- 経路制御プロトコルとの連携
  - Suffix配布済みのルータから経路制御プロトコルを起動
- ボーダルータとHANAサーバだけは設定の必要あり
  - ただしPrefix長の設定のみ、アドレス設定不要
  - ほとんどのノードは設定不要

## HANA のプロトタイプ実装

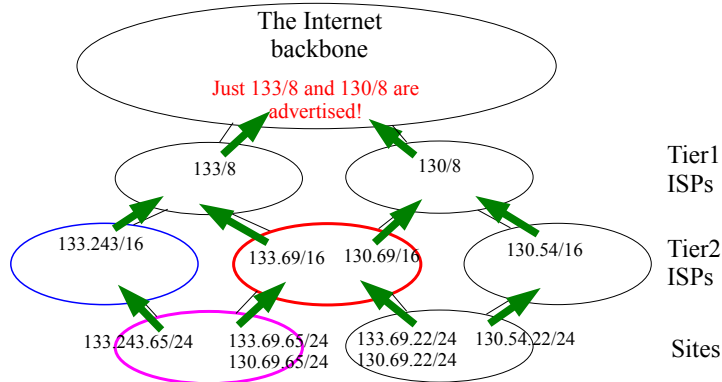
- Linux 上に HANA プロトコルのプロトタイプを実装
  - HANA プロトタイプは1プロセス上で複数のノードを動作させることができるシミュレータとして動作
  - サイトがマルチホームされたネットワークをシミュレータ上で構築
    - Suffix の自動割当及び、
    - マルチホームされている場合のPrefix の自動割当
- が正しく動作することを確認

## 基本となる要素ネットワーク

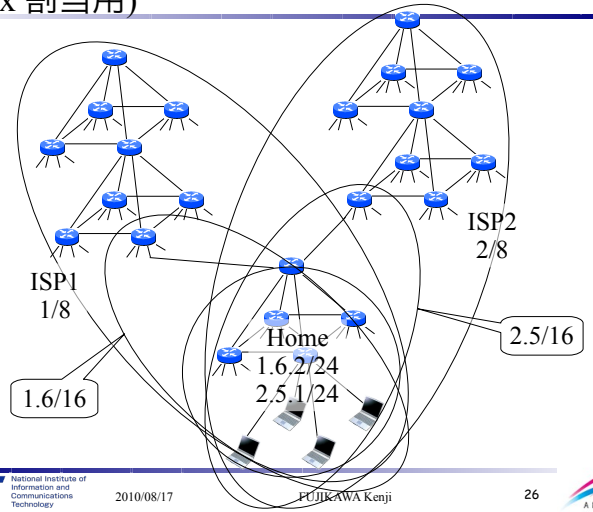


- R1やR4で上流と接続
- 図では省略されているが、全部で  $1+4+16=21$
- 要素ネットワークは /24 の空間で構成

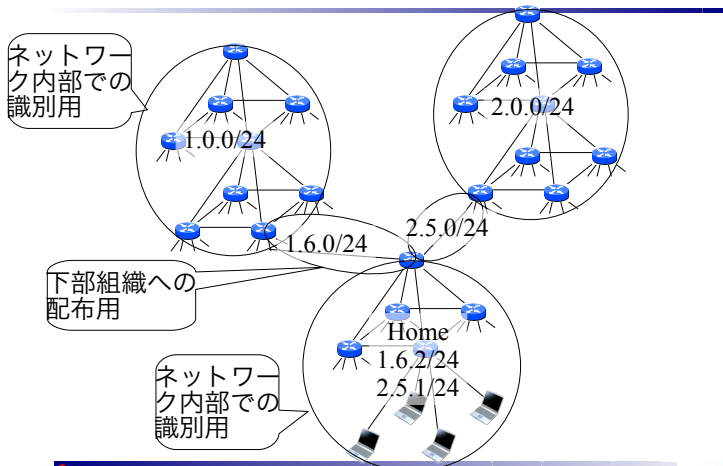
## マルチホームにおける階層的なアドレス割当てと広告



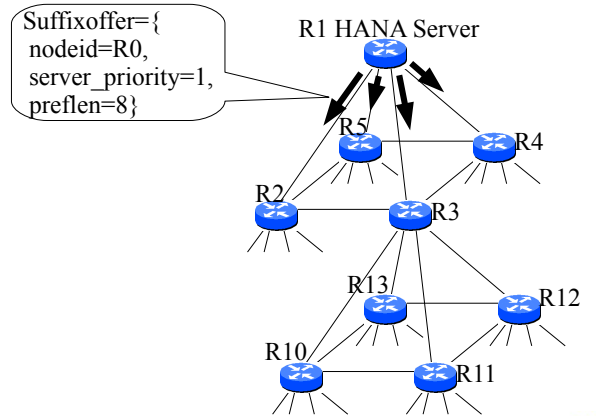
## HOME 及び ISP1、ISP2 のアドレス空間 (Prefix 割当て用)



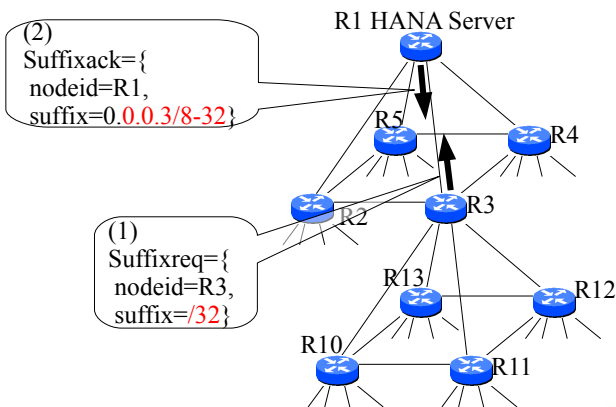
## Suffix 割当て用のアドレス空間



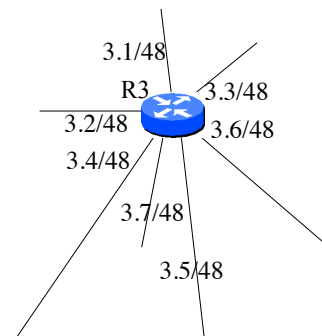
## HANA サーバ存在の通知



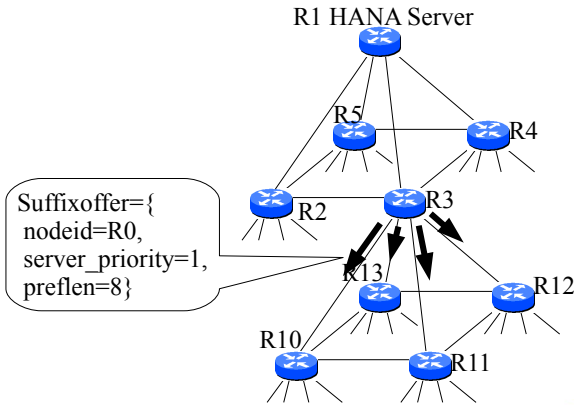
## Suffix 要求・応答メッセージ



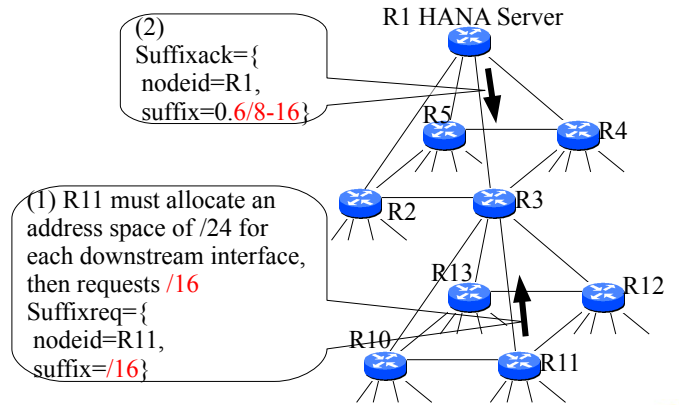
## インタフェースへのアドレス割当て



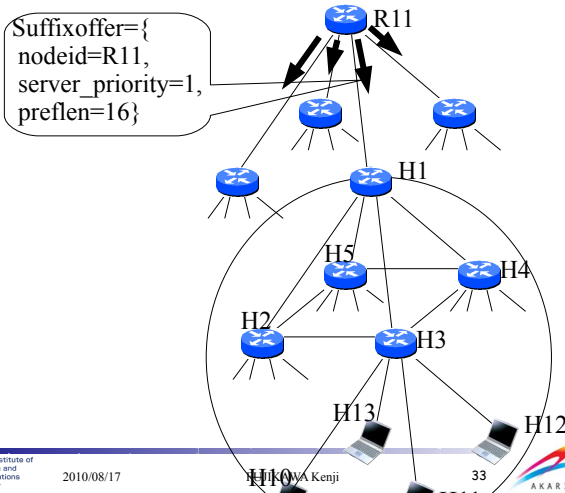
## R3 による HANA サーバ存在通知の転送



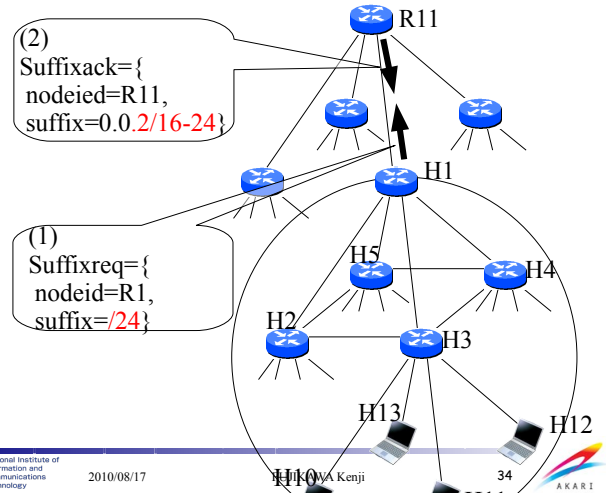
## 直接 HANA サーバに接続されていないルータのアドレス割当



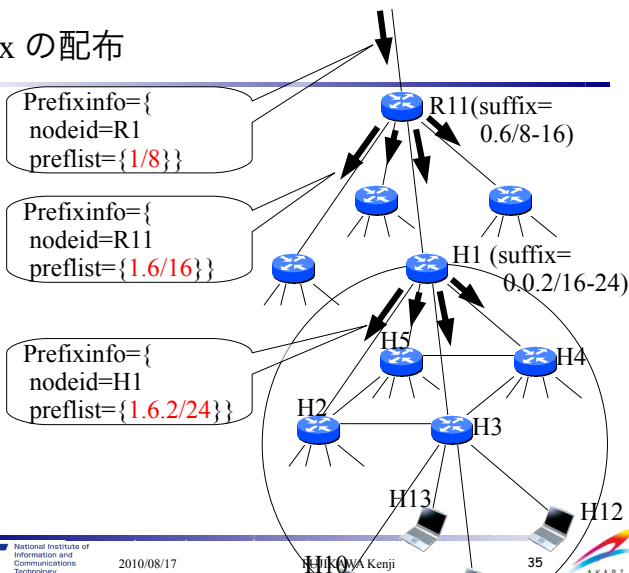
## 下流サイトへのアドレス割当(1)



## 下流サイトへのアドレス割当(2)



## Prefix の配布



## まとめ

- 研究の背景
  - End-to-endマルチホームはサイトに冗長性を提供しつつ経路表を削減できる
  - ルータの設定は簡単であるべき
- HANAプロトコルの提案とプロトタイプ実装
  - Prefixは上流から階層的に通知される
  - SuffixはDHCPのような方法で自動設定される
  - Suffix決定済ルータから経路制御プロトコル起動
- 今後の課題
  - HANAプロトコルのPCルータ上での稼働と評価
  - DNSデータベースの更新機構の設計