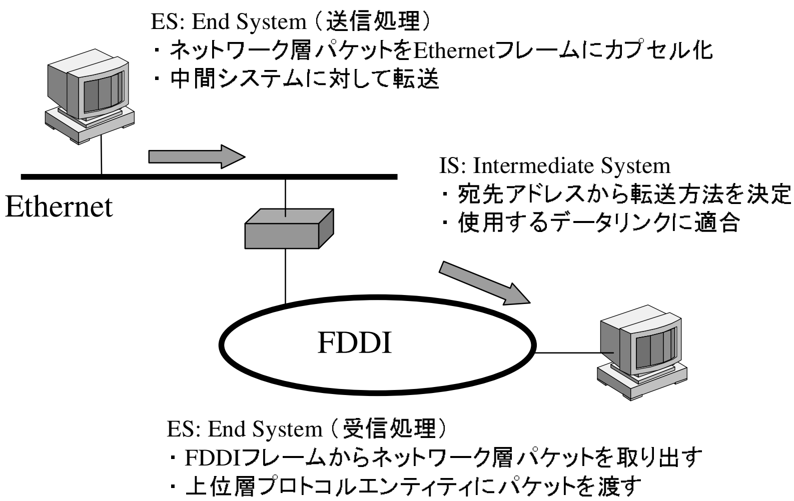
ネットワーク層プロトコルとIP

1999.5.12

ネットワーク層の役割

- ネットワーク内の任意のホスト間の通信を実現
  - » 複数のデータリンクをサポート
    - ネットワーク層パケットを、データリンクフレームにカプセル化
    - データリンクの違いを吸収
    - 異なるデータリンク間の相互接続
  - » 機能的な分類
    - 中間システム (IS: intermediate system)
      - 経路制御
      - データリンクに対する適合
    - 終端システム (ES: end system)
      - カプセル化
      - 経路制御

ISとES



実装モデル

- Connectionless
  - » 信頼性を保証しない
  - » best effort型
  - » e.g. IP (Internet Protocol), XNS
- Connection Oriented
  - » 信頼性の保証
  - » コネクション毎の設定可能
  - » e.g. X.25

基本機能

- 対象
  - » データリンク層から渡される受信したデータ
  - » トランスポート層から渡される送信データ
- 処理
  - » 経路制御
    - 受信したデータを転送するのか
    - どのデータリンクを利用するのか
    - 自ホストで受信するのか
  - » データリンクへの適合
    - 特に MTU (Maximum Transmission Unit) への適合
    - fragmentation / reassemble

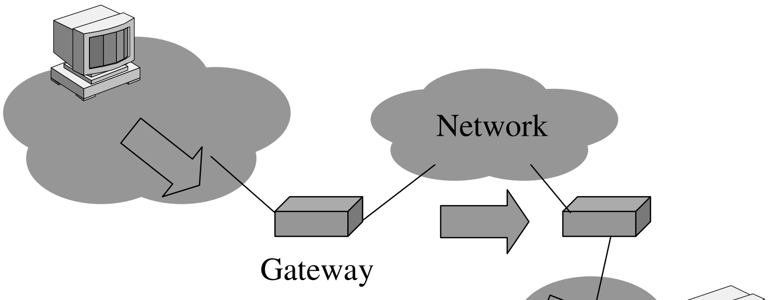
基本機能

- アドレスのマッピング
  - » ネットワーク層でアドレス空間を定義
  - » データリンク層とは独立
  - » 実際の通信処理では特定のデータリンク層を利用
  - » ネットワーク層アドレスとデータリンク層アドレスのマッピング
  - » N(a) → N(b)
    - N(a)→DL(a)（データ転送）DL(b) → N(b)
  - » ARP (Address Resolution Protocol)

IP

- IPv4: Internet Protocol (version 4)
- TCP/IPプロトコル群でのネットワーク層プロトコル
- 特徴
  - connectionless (IP datagram)
  - 4 octet address space with its own structure
  - unicast / multicast / broadcast / anycast
  - IP + ICMP + ARP + 経路制御
  - reassemble is done only at the end system
  - fragmentation is done based on MTU

モデル

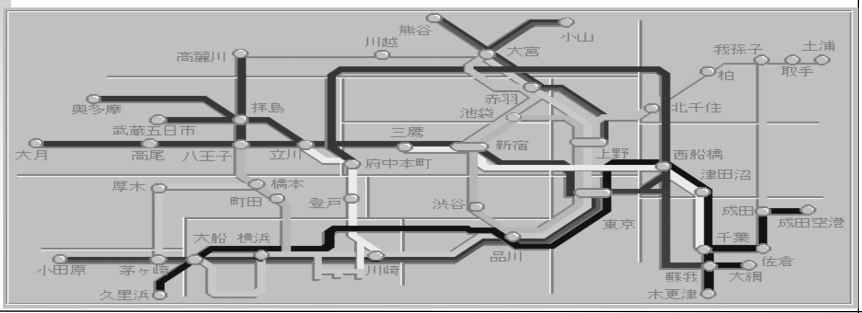


Gatewayが中間システムとして
IPデータグラムを経路制御機構
にしたがって転送

同一ネットワーク内はデータリンク
による直接通信

インターネットと鉄道網

- 客(データ)は必ず目的駅(コンピュータ)に到達する
- 駅には乗換え駅(ゲートウェイ)と普通の駅がある
- 各路線(ネットワーク)は自律運用、かつ、相互協調
- 選択可能な経路、経路の選択は知的な作業
- 鉄道網の「オーナー」はいない

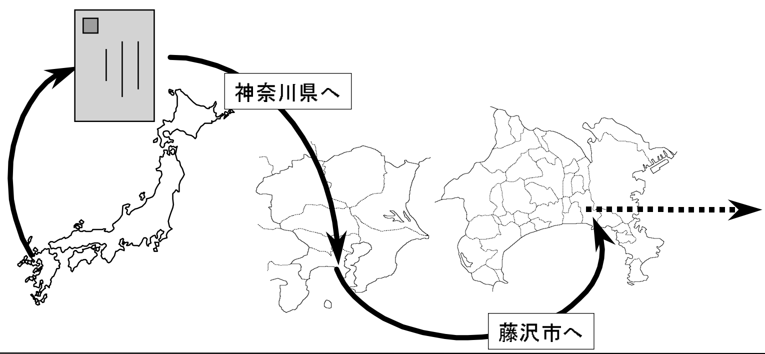


「東京乗り換え案内」より

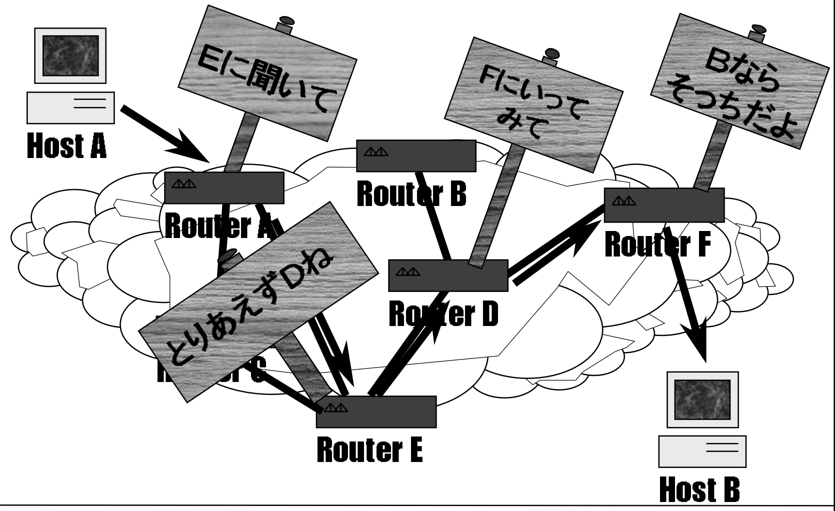
インターネットと郵便配送システム

- おおまかに次の配送先を決める
- 全ての郵便局が目的地を知っているわけではない
- 世界規模での自律分散協調システム

神奈川県藤沢市遠藤5322宛



Routerの役割



IPアドレス

- 4オクテット
- 構造を持つ
  - » ネットワーク部
    - ネットワーク毎に割り当てる
    - マスクを使ってネットワーク部を表現
  - » ホスト部
    - ネットワーク内で重複が無いように割り当てる
- ネットワークインタフェースに与える
  - » ゲートウェイでは複数のアドレスを持つ
- 1.0.0.0 から 223.255.255.255 まで

IPアドレス

163	221	74	127
0xA3	0xDD	0x4A	0x7F

ネットワーク部は24ビット

163.221.74.127/24

IPアドレス

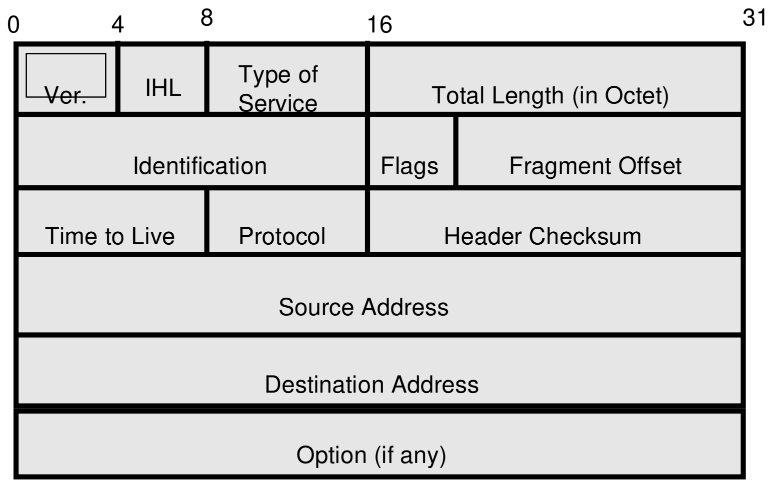
- ブロードキャストアドレス
  - ホスト部が all 1 の場合、そのネットワークに対するブロードキャスト (broadcast) となる
  - 例えば 163.221.74.255
  - そのネットワークに接続されており、稼動しているホストに対してデータグラムを配送

Copyright ©1999, Suguru Yamaguchi, All right reserved

IPアドレス

- ■ マルチキャストアドレス
  - » 224.0.0.0 から 240.255.255.255
  - » グループ通信
    - マルチキャストアドレスにデータを送信すると、「特定の」ホストのグループにデータを配送
    - グループは動的に形成
    - グループ管理

IPv4 header



Copyright ©1999, Suguru Yamaguchi, All right reserved

IPv4 Header

- Version
- Internet Header Length
- **Type Of Service**
- Total Length
- Identification
- **Flags**
- **Fragment Offset**
- **Time To Live**
- **Protocol**
- Header Checksum
- ***Source Address***
- ***Destination Address***
- IP options

IP Header

- Version
  - » 現在のプロトコルバージョン
  - » 4 (IPv4)
- IHL
  - » Internet Header Length
  - » 32bit word で数えたヘッダ長 (word数)
  - » IPオプションの部分が可変長であるために、必要なフィールド

Type Of Service

- 上位プロトコルによってつけられるタグ
  - » サービスを区別して処理したい場合に重要な情報が収められる
  - » IPデータグラムの処理についての「希望・お願い」
- フォーマット

0 7

Precedence	D	T	R	未使用

Type of Service

- Precedence
  - » 処理優先権
  - » 0から7の値
  - » 各データグラムの重要度を表す
- 残り3ビットは転送処理に対する希望を表す
  - » ビットが設定されているものを希望
  - » D: low Delay
  - » T: high Throughput
  - » R: high Reliability

Type of Service

- 幾つかのアプリケーションの実装では設定
- しかしながら基本的に企画倒れ
  - » 処理優先権の考え方
  - » プライオリティと網管理

Total Length

- IP データグラムの全長
- オクテット単位
- ヘッダ＋データ

Time To Live

- ネットワーク内でのパケット生存時間
  - 迷子のIPデータグラムを消滅させる
  - 設計時には具体的な時間の情報を考えていた
  - 時間の表現は難しい
- 最大ゲートウェイホップ数
  - ゲートウェイ通過毎に値を減らす
  - 0になったらデータグラムを捨てる
  - 送信ホストに通知 (ICMP Time Exceeded)

Fragmentation & Reassemble

- データリンク毎にMTUは決まっている
  - データリンクフレームのデータ部分の最大長
  - 例えば Ethernet であれば 1500オクテット
- IPデータグラム
  - 可変長
  - 最大長 64Kbyte (Total Length Fieldの制限)
- 大きなデータをMTU以下に分割
  - fragmentation
  - ゲートウェイ毎に発生する可能性あり

Fragmentation & Reassemble

- 最終的な受信ホストで再構成
  - » reassemble
  - » 中間システムでは reassemble してはいけない
  - » 性能の劇的な悪化をさける

Fragmentation & Reassemble

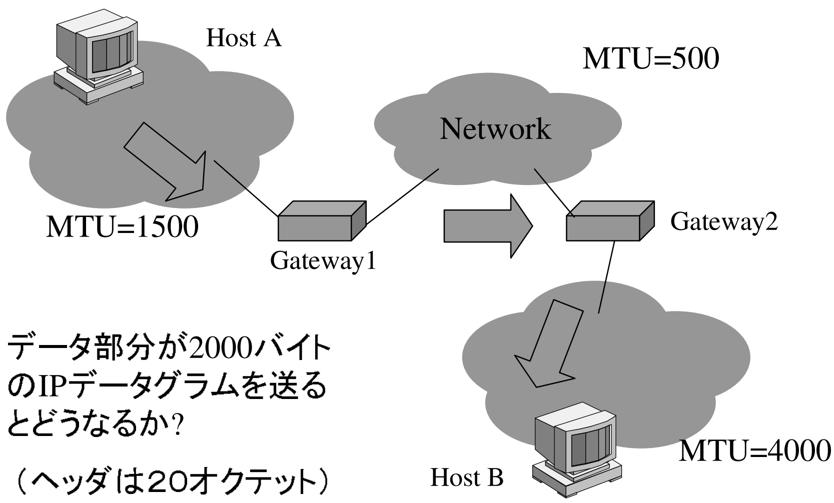
- Flag フィールド（3ビット）
  - » 0, DF, MF
  - » DF: Don't Fragment
    - このビットが設定されているデータグラムは fragment 処理をしてはならない
    - MTUよりもIPデータグラムが大きい場合にはエラー
  - » MF: More Fragment
    - フラグメントされていて、かつ、後続の fragment がある場合に設定

Copyright ©1999, Suguru Yamaguchi, All right reserved

Fragmentation & Reassemble

- Fragment Offset (13bit)
  - » そのデータグラムが元のメッセージのどの部分か
  - » MFと Fragment Offset を元に Reassemble を実行

Fragmentation & Reassemble



データ部分が2000バイト
のIPデータグラムを送る
とどうなるか?

（ヘッダは20オクテット）

Fragmentation & Reassemble

- Host A
  - » H+1480: MF=1, FO=0
  - » H+520: MF=0, FO=1480
- Gateway 1
  - » H+480: MF=1, FO=0
  - » H+480: MF=1, FO=480
  - » H+480: MF=1, FO=960
  - » H+40: MF=1, FO=1440
  - » H+480: MF=1, FO=1480
  - » H+40: MF=0, FO=1960
- fragment が一つでもなくなれば、IPデータグラム全体を棄却
  - » 性能劣化につながる
  - » NFSでは問題

上位プロトコルとの関係

- Protocol フィールド
  - » 上位プロトコルの番号を指し示す
  - » multiplexing/de-multiplexing の処理
  - » TCP (6), UDP (17)
  - » プロトコルスイッチによる上位層処理の駆動

IPトンネリング

- IPデータグラムをIPデータグラムで運ぶ
- Protocol フィールドには特別な値
  - » IP over IP
- トンネリング (tunneling) は最近では広く使われている。
  - » IPsec / VPN (Virtual Private Network)
  - » IP Multicasting
  - » 仮想ネットワーク (Mbone, 6bone)

Network Byte Order

- IPヘッダの数値フィールドの符号化方法
- RFC791で規定
- 各ホストのバイトオーダとは独立に定義
  - » 解釈に違いが出ないようにするため
- Big-Endian
  - » MSBが手前
  - » パケットフォーマットで表記すると左側

Checksum

- 16bit 1's complement arithmetic
- ヘッダのチェックサム
  - » 計算する場合には、checksum フィールドは all 0 であることを仮定して計算
- 計算方法
  - » ヘッダを16ビット毎に分割し、1の補数を計算
  - » 和計算
  - » 結果の1の補数
- 受信側ではヘッダの checksum を再計算・比較

アドレス解決 (address resolution)

1999.5.12

Address Resolution

- ネットワーク層アドレスとデータリンク層アドレスのマッピング
  - » 経路制御の役割
    - データリンク層の選択
    - 次に送るべきホストの選択
  - » 実際の転送処理で必要となる相手のデータリンク層アドレスの獲得

いろいろな方式

- テーブル方式
  - » 各ホストで対応表を管理する
  - » 静的な設定
  - » 規模の小さなネットワークでは有効
  - » ホストのネットワークインタフェースが変更された場合には、全てのホストでテーブルの更新が必要

いろいろな方式

- サーバ方式
  - » テーブル方式の拡張
  - » サーバで情報を一括管理
  - » サーバとの通信をどのように実現するのか
    - well-known アドレスを利用
    - デフォルトアドレスを利用
    - broadcast (データリンクレベル)を利用
  - » サーバがダウンした場合の対処方法はどうするのか
    - サーバの二重化？

いろいろな方法

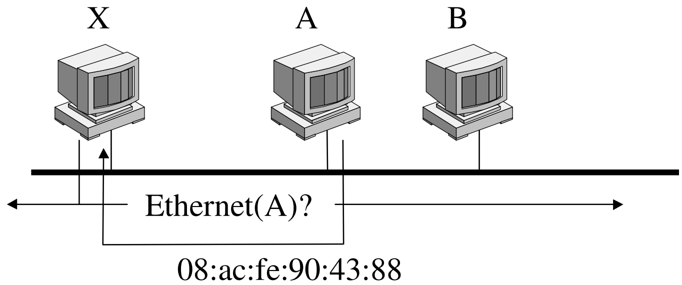
- 動的決定方法
  - » ARP (Address Resolution Protocol) の利用
  - » 通信時毎に決定
    - Dynamic Binding
  - » ネットワーク層に依存しない通信方法が必要

ARP

■ RFC826

■ データリンクレベルでの同報通信を用いて問い合わせを実行

■ 問い合わせを受けたホストがアドレス情報を返送



問題点 (1)

■ データリンク層毎にARPの方式を決定する必要がある

» RFC826 は汎用のARPを定めている

- Broadcast Channel の存在
- 送信側データリンクアドレスがフレームの中に記録
- ローカルのアドレス情報が決定可能

» この範疇をはみ出るデータリンクでは利用できない

問題点 (2)

- ブロードキャストアドレスを返送するクライアントがある場合にはどうするのか
  - » Ethernet Meltdown….
  - » 明らかに実装上の不備
- 「うそつき」がいた場合にはどうするのか
  - » 勝手に他のホストのアドレスを答える
  - » 通信出来なくなる可能性が高い
  - » セキュリティ問題

UNIXでの実装

- テーブル方式と動的決定方式の両方を採用
  - » arp コマンドによるテーブルエントリの作成・削除
  - » ARPプロトコルによる動的決定
  - » 問い合わせ回数を減らすために、ARPによる動的決定された情報はキャッシュされる
    - 一定時間内はそのキャッシュを利用
    - 一定時間後は改めて問い合わせ
    - 実用的な処理