ネットワーク構築概論
～配線から経路制御まで～

2002年12月17日

株式会社インターネットイニシアティブ

山口 二郎 (jiro-y@iij.ad.jp)

IIJ Internet Initiative Japan

目的

- データリンク層とネットワーク層の役割は
- 障害が起こりにくいネットワークを設計するには
- ネットワークの冗長化を行うには
- ルーティングとは

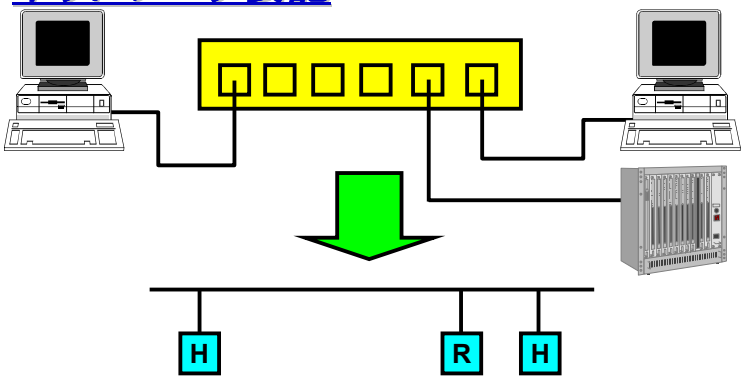
IIJ Internet Initiative Japan

発表内容

- データリンク層とネットワーク層の役割
- ハブ、スイッチ、ルータの違い
- ネットワーク設計
- アドレスの割り当てポリシー
- ネットワークの冗長化
- ダイナミックルーティングの動作原理
- ダイナミックルーティングを用いたバックアップ、バランシング

IIJ Internet Initiative Japan

ネットワーク表記



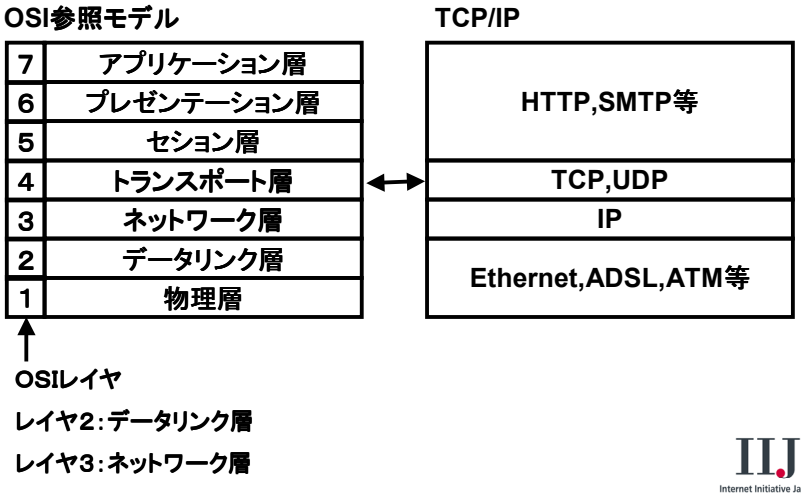
- ハブ、スイッチなどは1本の線またはSWで表わします。
- ホストはH、A、B、C、D等で、ルータはR等で表記します
- レイヤ3スイッチなどは説明中ではルータと区別していません

IIJ Internet Initiative Japan

データリンクフレームとルーティング

- ここではデータリンク層とネットワーク層の役割を解説します
- MACアドレス（イーサネットアドレス）とIPアドレスの両方のアドレスが必要な訳
- データリンク層の種類
- ルーティングがなぜ必要なのか
- ルーティングがなくても通信できるのはなぜか

OSI参照モデルとTCP/IP

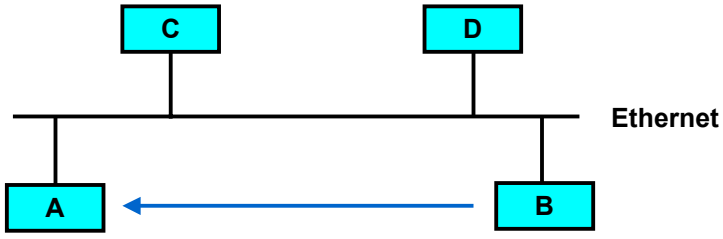


データリンク層の種類

- Multi Access Media (ARP)
  - MAC(Media Access Control)アドレスを用いて通信を行う
  - MACアドレスとIPアドレスとの対応はARP(Address Resolution Protocol)を用いる
  - Ethernet等
- Multi Access Media (固定)
  - 特定の識別子とIPアドレスに結び付け、固定的に設定を行う
  - フレームリレー、ATM等のMulti Access Mode
  - EthernetでIPアドレスとMACアドレスを固定的に設定
- Point to Point Media
  - 通信相手が物理もしくは仮想I/Fで特定されるもの
  - 64k,128k,1.5M,6M,45M,150M,600M,2.4G,10Gなどの専用線
  - フレームリレー、ATM等のPoint to Point Mode

IIJ Internet Initiative Japan

ARPの動作ー1



Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	不明
C	192.168.0.3	不明
D	192.168.0.4	不明

Host AのIP/MACアドレス対応表

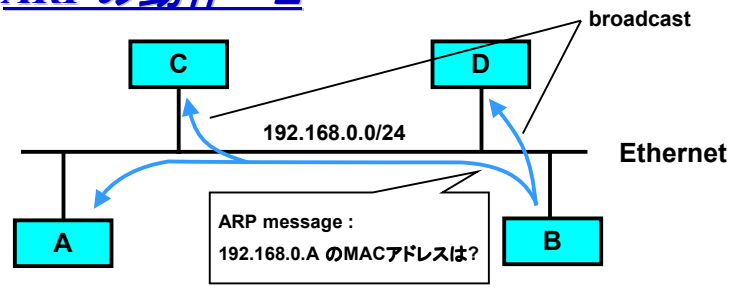
Host	IPアドレス	MACアドレス
A	192.168.0.1	不明
B	192.168.0.2	00-00-f8-04-c0-11
C	192.168.0.3	不明
D	192.168.0.4	不明

Host BのIP/MACアドレス対応表

- BはAに通信したいが、BはAのMACアドレスがわからない

IIJ Internet Initiative Japan

ARPの動作ー2



Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	不明
C	192.168.0.3	不明
D	192.168.0.4	不明

Host AのIP/MACアドレス対応表

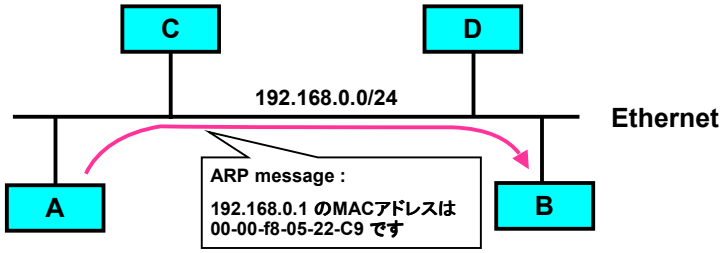
Host	IPアドレス	MACアドレス
A	192.168.0.1	不明
B	192.168.0.2	00-00-f8-04-c0-11
C	192.168.0.3	不明
D	192.168.0.4	不明

Host BのIP/MACアドレス対応表

- BはAのMACアドレスを尋ねるメッセージをbroadcastする

IIJ Internet Initiative Japan

ARPの動作ー3



Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	不明
C	192.168.0.3	不明
D	192.168.0.4	不明

Host AのIP/MACアドレス対応表

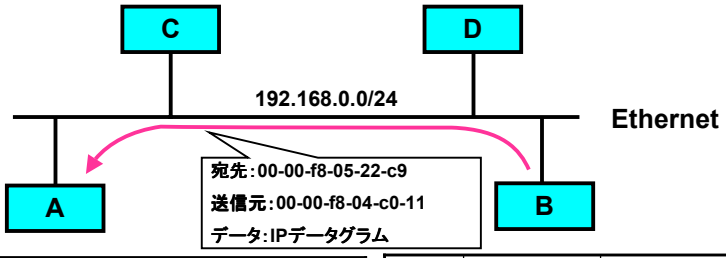
Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	00-00-f8-04-c0-11
C	192.168.0.3	不明
D	192.168.0.4	不明

Host BのIP/MACアドレス対応表

- Aは自分のMACアドレスをBに返答する

IIJ Internet Initiative Japan

ARPの動作ー4



Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	不明
C	192.168.0.3	不明
D	192.168.0.4	不明

Host AのIP/MACアドレス対応表

Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	00-00-f8-04-c0-11
C	192.168.0.3	不明
D	192.168.0.4	不明

Host BのIP/MACアドレス対応表

- BはAに対してデータを送ることができるようになる

Multi Access Media(ARP)ー1

- ARP(Address Resolution Protocol)
  - ARPとはIPアドレスとMACアドレスを対応させるためのプロトコル(IP以外のプロトコルでも利用されますが、IPに限って説明します)
- IP/MACアドレス表
  - IP/MACアドレス対応表のことを「ARPテーブル」「ARPキャッシュテーブル」「ARPキャッシュ」などと呼ばれている
- ARPキャッシュ
  - ARPテーブルに登録されたIP/MACアドレスは一定時間保持(キャッシュ)される
  - ARPテーブルにIP/MACアドレスが存在するときはARPによるbroadcastは行われずに、ARPテーブルにしたがって通信が行われる。
  - 一定時間後、IP/MACアドレスはARPテーブルから削除され、その後通信が行われた場合には再びARPを実施する
  - キャッシュすることで、ARPによるデータリンク層のbroadcastを抑制している

Multi Access Media(ARP)－2

- ARPキャッシュのクリア
  - 機器の交換などでIP/MACアドレス対応に変化がある場合はARPキャッシュをクリアを行う必要がある場合がある
    - arp –d (ホストなど)
    - clear arp (ルータなど)
  - 最近のネットワーク機器やOSは機器交換後に明示的にARPキャッシュをクリアしなくても高速にARPキャッシュ反映されるような実装が増えている

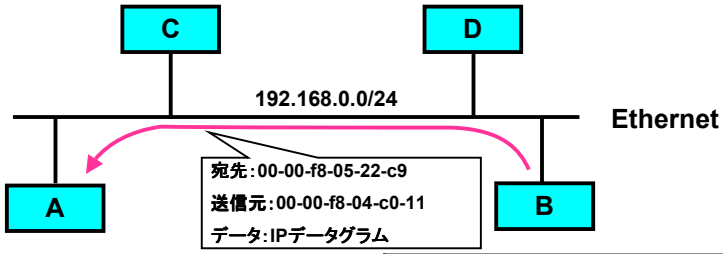
IIJ Internet Initiative Japan

Multi Access Media(ARP)－3

- ARPのメリット
  - 他の機器のIP/MACアドレス対応表を設定する必要が無い
  - 機器交換を行ってもARPキャッシュがクリアされれば自動的に反映される
- ARPの運用上の注意点
  - 機器交換の際にARPキャッシュをクリアしないとすぐに通信できないことがある
  - broadcastが利用されるため大規模なレイヤ2ネットワークでは帯域を圧迫する
  - Globalセグメントで多くの利用されていないアドレスが存在すると、インターネットから未使用アドレスに対するアクセスによりLANが輻輳することがある
    - インターネット上のウイルスに感染したホストなどからのポートスキャンにより発生する(NIMDAなど)
    - 未使用アドレスの個数×リトライ回数のbroadcastが発生する

IIJ Internet Initiative Japan

固定IP/MACアドレス対応表の動作



Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	00-00-f8-04-c0-11
C	192.168.0.3	00-60-08-07-07-ac
D	192.168.0.4	00-a0-24-4a-7a-12

Host AのIP/MACアドレス対応表(固定)

Host	IPアドレス	MACアドレス
A	192.168.0.1	00-00-f8-05-22-c9
B	192.168.0.2	00-00-f8-04-c0-11
C	192.168.0.3	00-60-08-07-07-ac
D	192.168.0.4	00-a0-24-4a-7a-12

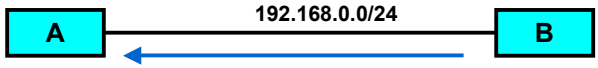
Host BのIP/MACアドレス対応表(固定)

- IP/MACアドレス対応表は事前に固定的に設定されるため、BはAに対してデータを送ることができる

Multi Access Media(固定)

- ARPを用いず固定的に物理アドレスとIPアドレスを結びつける
- ARPを用いないためbroadcastが発生しない
- broadcastが利用できないため、ARPが利用できない場合に利用
- 機器交換などでIP/MACアドレス対応が変化する場合にはすべての機器の設定を変更する必要がある
- ATMではVPI/VCIを固定的に設定する

Point to Point Mediaの動作



Host	IPアドレス
A	192.168.0.1
B	192.168.0.1以外の 192.168.0.0/24

Host Aの通信先

Host	IPアドレス
A	192.168.0.2以外の 192.168.0.0/24
B	192.168.0.2

Host Bの通信先

- Point to Point Mediaに属しているすべてのネットワークは相手側に送り出す（ARPや固定アドレス表は不要）
- Point to Point Mediaから来たフレームはすべて受け取る
- IP層によってはA、B間をループしてしまうこともある

IIJ Internet Initiative Japan

Point to Point Media-1

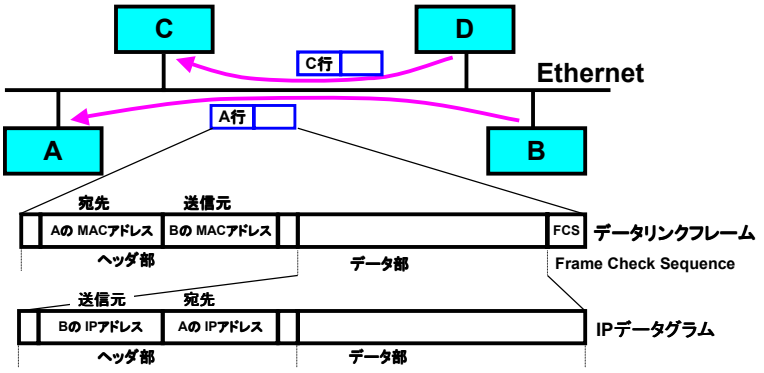
- 自分以外の属しているネットワークに対するすべての通信をPoint to Point Mediaに送り出す
- Point to Point Mediaから来たフレームはすべて受け取る
- 受け取ったフレームはIP層で評価される
- IP層の評価によってはPoint to Point Mediaでループすることもある
- すべてのフレームを選択せずに送り出し、受け取るためMACアドレス、broadcastは不要

IIJ Internet Initiative Japan

Point to Point Media-2

- ATM専用線も設定によりPoint to Point Mediaとして利用することが可能
- ネットワークは一般的に/30もしくはunnumberedが利用される
  - 192.168.0.0/30 (ネットワーク例)
    - 192.168.0.1 (Router 1)
    - 192.168.0.2 (Router 2)
  - unnumberedインターフェースへのルーティングはインターフェース名などが利用される
    - ip route 172.16.0.0 255.255.0.0 *Serial0/0*

Ethernetを流れるIPデータグラム



Connectedなネットワーク

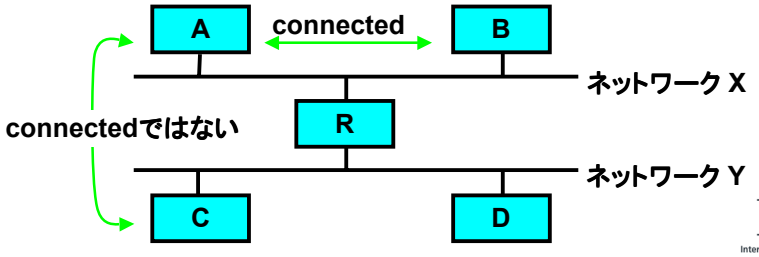
- A、Bは直接同じネットワークに接続している
  - MACアドレス、IPアドレスの対応表を ARP(address resolution protocol)などにより持っている

↓

- これを「connected」な状態という

↓

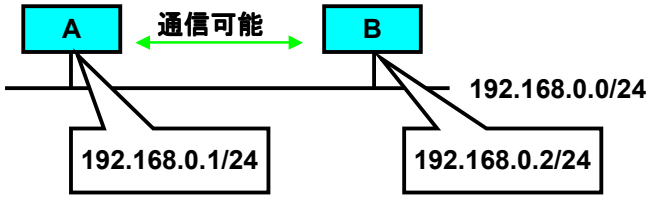
- ルーティング設定が不要で、ハブなどで接続すると通信できる



IIJ Internet Initiative Japan

ネットワーク層から見たConnectedなネットワーク－1

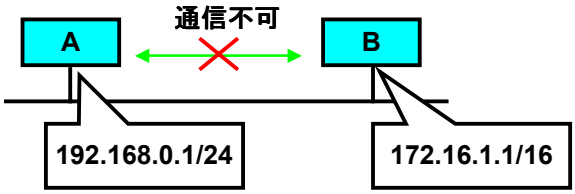
- Aのアドレス
  - 192.168.0.1/24
- Aから見たConnectedなアドレス空間
  - 192.168.0.0 ～ 192.168.0.255
- Bに192.168.0.2 ～ 192.168.0.254のアドレスを付ける
  - Bに192.168.0.2を付ける
  - A-B間の通信が可能



IIJ Internet Initiative Japan

ネットワーク層から見たConnectedなネットワーク－2

- Aのアドレス
  - 192.168.0.1/24
- Aから見たConnectedなアドレス空間
  - 192.168.0.0 ～ 192.168.0.255
- Bに192.168.0.2 ～ 192.168.0.254以外のアドレスを付ける
  - A-B間の通信ができない



Connectedではないネットワーク－1

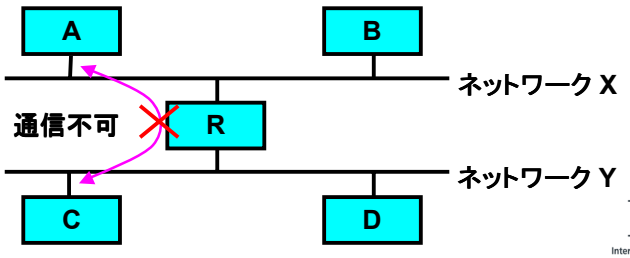
- A、Cはそれぞれ異なるネットワークに接続しているため connectedではない
- ルーティング*設定なし*ではA、C間の通信はできない

Aのルーティングテーブル

destination	Next Hop	到達性
X	Connected	到達可
Y	なし	到達不可

Cのルーティングテーブル

destination	Next Hop	到達性
X	なし	到達不可
Y	Connected	到達可



Connectedではないネットワーク－2

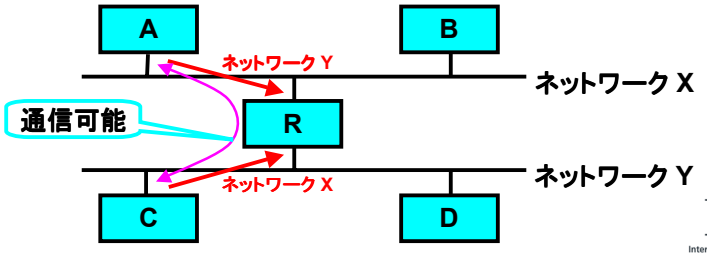
- ルーティング設定を行なう
  - A: ネットワークYを Rにルーティング
  - C: ネットワークXを Rにルーティング
- これにより、A⇔C間の相互通信が可能となる
  - Rは A,C共に connectedなため、アドレスを設定するだけで通信が可能

Aのルーティングテーブル

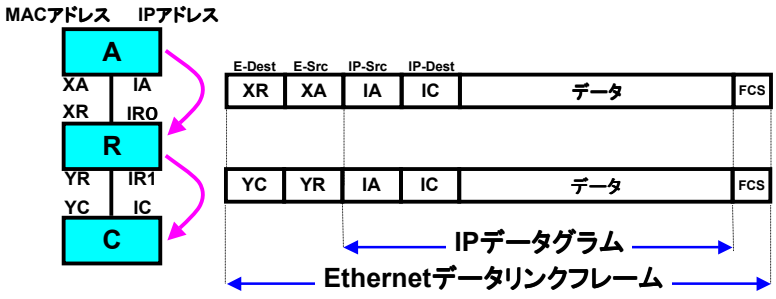
destination	Next Hop	到達性
X	Connected	到達可
Y	R	到達可

Cのルーティングテーブル

destination	Next Hop	到達性
X	R	到達可
Y	Connected	到達可



データリンクフレームの状態



- IPデータグラムの宛先、送信元は途中で変化しない
- データリンクフレームはルータを通過する毎に変化する
- 「データリンクフレームの宛先」＝「IPデータグラムの宛先」とは限らない

ネットワーク用語のまとめ

- Destination、Destination Address
  - 目的地という意味、ネットワークでは文字どおり目適地アドレス、宛先アドレスとして扱われる。Destination（デスティネーション）とそのまま使われることが多い。経路制御ではアドレスだけでなくマスク情報を含んだネットワーク情報もDestinationとして扱われる。
- NEXT HOP、NEXT HOP Address
  - 次に配送すべきアドレス。ルータやホストはDestinationがConnectedでない場合に次に配送すべきアドレス（NEXT HOP）を参照してIPパケットを送信する。IPパケットを受け取ったルータやホストはその次に配送すべきアドレス(NEXT HOP)に送信し、これらを繰り返してDestinationに到達する。
- ルーティング、ルーティング情報
  - 経路。DestinationとNEXT HOPをペアとしたもの。
- ルーティングテーブル
  - ルータやホストが持っているルーティングの一覧
- ルーティングする
  - ルータが正常にルーティングテーブルに基づいてIPパケットを送り出している状態「このルータはきちんとルーティングしている」

IIJ Internet Initiative Japan

データリンクフレームとルーティングのまとめ

- データリンク層、ネットワーク層共にConnectedな状態であればルーティング設定をせずに通信が可能
- Connectedでないネットワーク、ホストとの通信には必ずルータの設置、ルーティング設定が必要
- IPデータグラムの宛先、送信元は途中で変化しない
- データリンクフレームはルータを通過する毎に変化する
- 「データリンクフレームの宛先」＝「IPデータグラムの宛先」とは限らない

IIJ Internet Initiative Japan

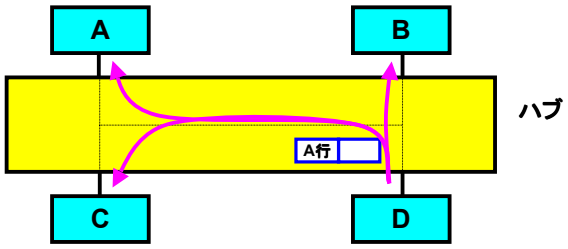
スイッチとルータの機能の違い

- ハブとスイッチの機能の違い
- スイッチを有効に使う方法
- ルータを利用するための設定
- ネットワーク設定の自動化
- スイッチとルータの違い
- スイッチの耐障害性
- ルータの耐障害性
- Broadcast flood問題

IIJ Internet Initiative Japan

ハブとスイッチの違い-1

ハブで構成した場合

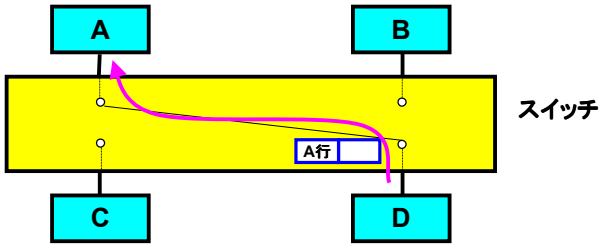


- ハブは全てのポートが常時接続された状態になっている
- このため異なるポート間の通信を、通信に関係の無い他のポートに伝播して、他の通信を妨げる

IIJ Internet Initiative Japan

ハブとスイッチの違い-2

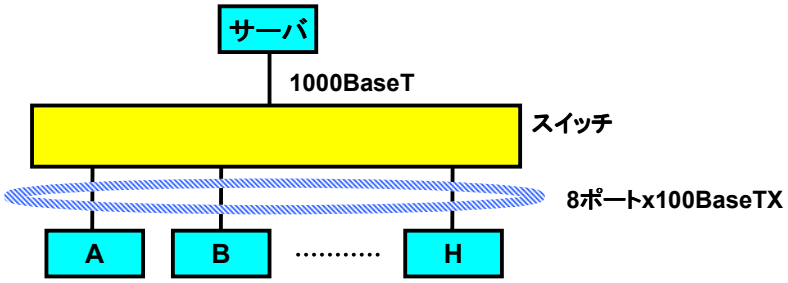
スイッチで構成した場合



- スイッチは、ポート毎に接続されている機器のMACアドレスを学習し、通信時には必要なポート間のみで通信する

IIJ Internet Initiative Japan

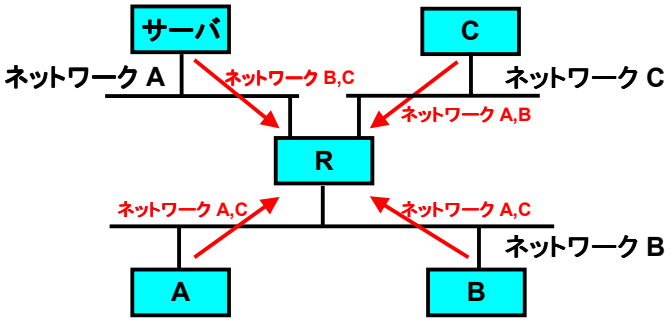
スイッチを有効に使うには



- 主にサーバ、ホスト間のトラフィックの場合に有効
- A ⇔ サーバ
⋮
H ⇔ サーバ } それぞれ100BaseTXをフルに利用可能

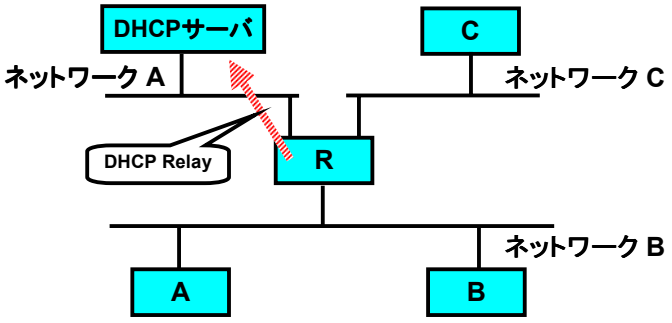
IIJ Internet Initiative Japan

ルータを利用するための設定－1



- ネットワークをサブネットに分割する
- 通信相手のネットワークのルーティングを設定する
  - DHCP,ダイナミックルーティングプロトコルなどで自動化することもできる

ルータを利用するための設定－2



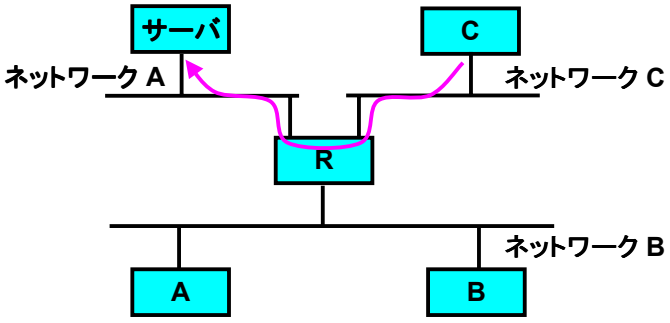
- DHCPサーバ設定
  - DHCPサーバは同一ネットワークに存在する必要がある
  - ルータのDHCP Relay設定により異なるネットワークでもDHCPを利用できる

ネットワーク設定の自動化

- DHCP（Dynamic Host Configuration Protocol）
  - アドレスの自動割り当てを行う
  - RFC2131
  - 主にクライアントで用いられる
  - Renumberを自動的に行うため、ポータビリティがある
- ダイナミックルーティングプロトコル
  - 自動的にルーティングが設定される
  - 主にルータ間で用いられる
  - RIP,RIP2,OSPFなどがある
  - 障害時に迂回路などを自動的に選択する

IIJ Internet Initiative Japan

スイッチとルータの違い



- ルータは、あるネットワーク間の通信を他の関係の無いネットワークに伝播しない

IIJ Internet Initiative Japan

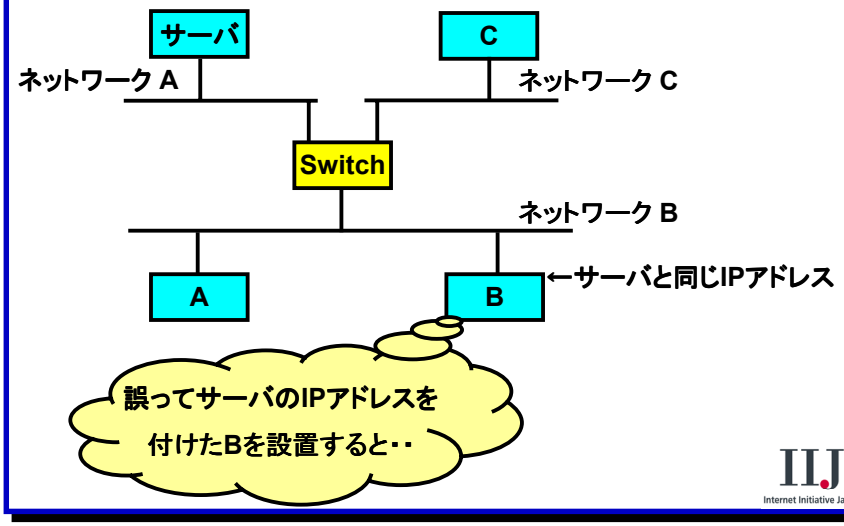
スイッチとルータの機能の違い

- ハブとスイッチの機能の違い
  - スイッチは異なるポートの通信を他のポートに伝播しない
- スイッチとルータの違い
  - ルータは異なるネットワークの通信を他のネットワークに伝播しない
  - スイッチとは異なり、ルーティングの設定が必要
  - サブネット分割が必要
- スイッチを有効に使うには
  - トラフィックが集中するようなポートにはスイッチを導入する

次に問題点について検討する

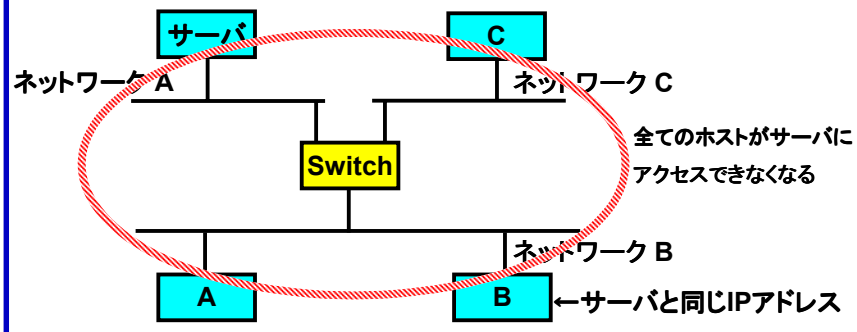
IIJ Internet Initiative Japan

スイッチの耐障害性-1



IIJ Internet Initiative Japan

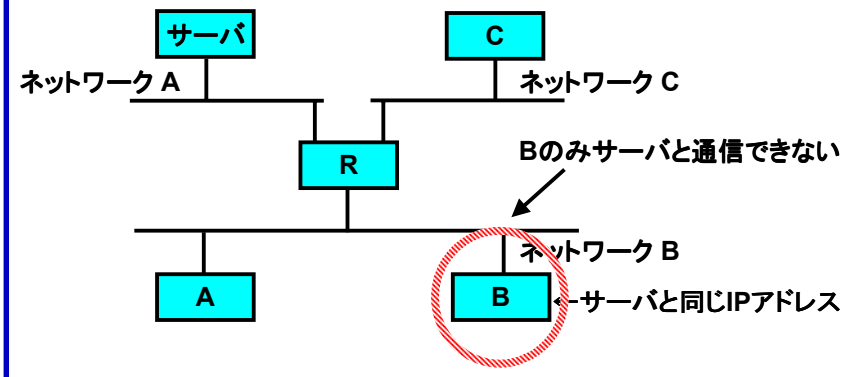
スイッチの耐障害性-2



- スイッチでは、1クライアントの間違った設定の影響がネットワーク全体に及ぶ

IIJ Internet Initiative Japan

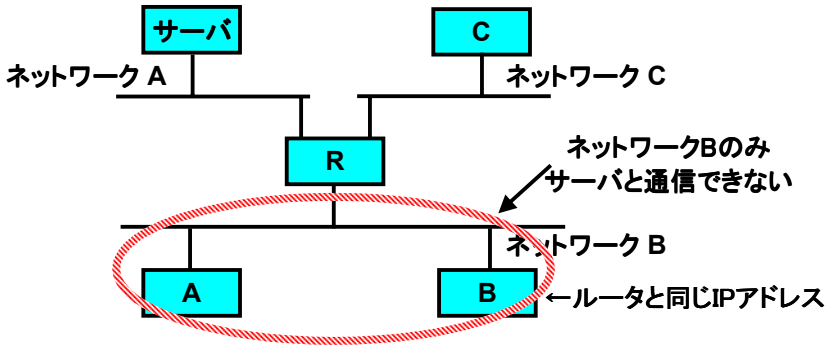
ルータの耐障害性-1



- ルータでは、1クライアントの間違った設定があったとしても、ネットワーク全体に影響を与えることはない

IIJ Internet Initiative Japan

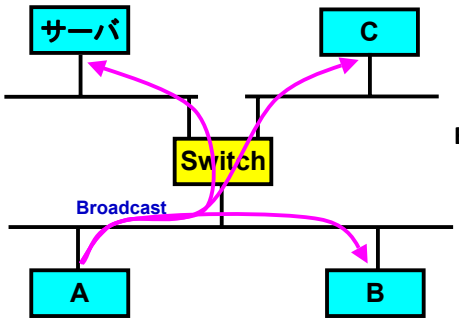
ルータの耐障害性-2



- 最悪の場合でも、ルータでは1クライアントの間違った設定の影響は同一セグメント内にとどまる

IIJ Internet Initiative Japan

Broadcast Flood-1

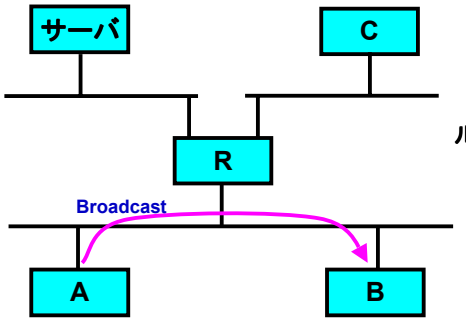


Broadcastパケットはスイッチの全てのポートに影響を与える

- ホスト数が増えると、broadcastパケットも無視できないトラフィックとなる
- Windows系の OSはこのようなbroadcastパケットを大量に発生させる傾向がある

IIJ Internet Initiative Japan

Broadcast Flood-2



ルータは Broadcastパケットを
他のネットワークに通さない

- Broadcast floodは発生しない
- 大規模ネットワークにも対応

IIJ Internet Initiative Japan

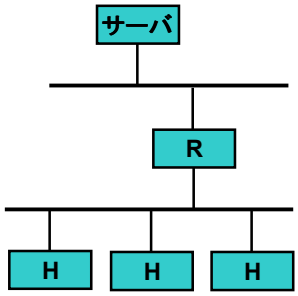
スイッチ VS ルータ

- スイッチの利点
  - ルーティングを考慮しなくて良い
  - ハブに比べて効率的なネットワークを構築することができる
- ルータの利点
  - ダイナミックルーティングプロトコルでバックアップ構成が可能
  - Broadcast floodが発生しない
  - 規模が大きくなってもスケールする
  - 障害時に被害を最小限に抑えることができる
  - 障害時の切り分け作業が比較的行いやすい
- 結論
  - ルータでサブネット化を行い、トラフィックが集中するようなポートにはスイッチを導入する

IIJ Internet Initiative Japan

ネットワーク設計-1

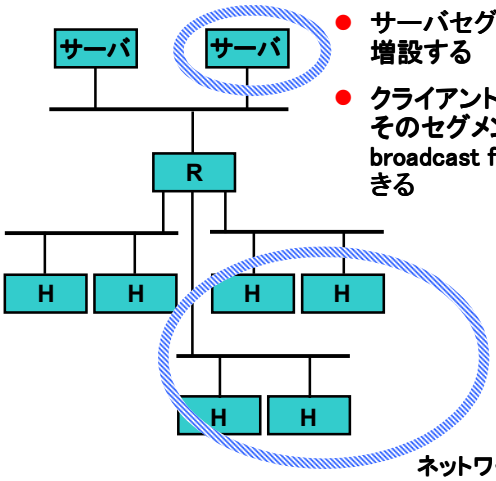
左図ネットワーク構成の特徴



- 小規模であってもサーバのセグメントを分離する
  →サーバの安全性を確保する
- クライアントはDHCPによりアドレスの割り当てとdefault経路を得る
- Broadcast floodのサーバへの影響を防ぐ

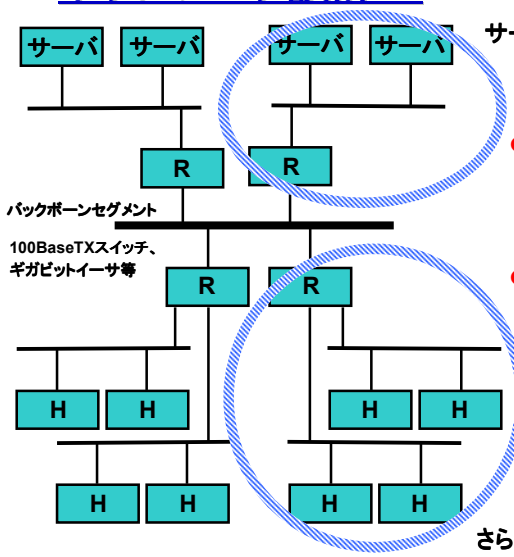
ネットワーク設計-2

サーバの増設



- サーバセグメントの安全性を保ちつつ増設する
- クライアントセグメントのbroadcast をそのセグメント内に留められるため broadcast flood現象の発生を抑制できる

ネットワーク設計-3



- さらにセグメントを増設する際は、バックボーンセグメントを高速化するだけで、対応が可能
- スイッチングベースのネットワークから、左図のような大規模ネットワークに拡張する場合、リナンバリングが避けられない

IIJ Internet Initiative Japan

ネットワーク設計のまとめ

- スケーラビリティを考慮するとサブネット化は不可欠
- 安全性を考慮してサーバは別のセグメントに
- トラフィックの集中するサーバ、ルータなどにはスイッチを導入する
- 規模の拡大を見越したネットワークトポロジの設計

↓

ネットワーク規模拡大を考慮したアドレス割り当て

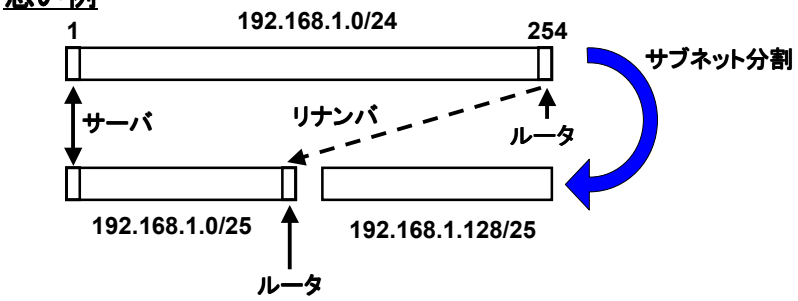
IIJ Internet Initiative Japan

アドレスの割り当てポリシーとは

- 規模の拡大を想定しネットワークアドレスの組織内割り当てを考える
- アドレスを先頭から詰めて使用するべきか、それとも先頭と後ろから使用していくべきか
- 各部署に割り当てる時は、どのように割り当てていけばいいのか
- 各部署内で各ホストに割り当てる場合にどのように割り当てていけば良いのか

組織全体でのアドレスの割り当て-1

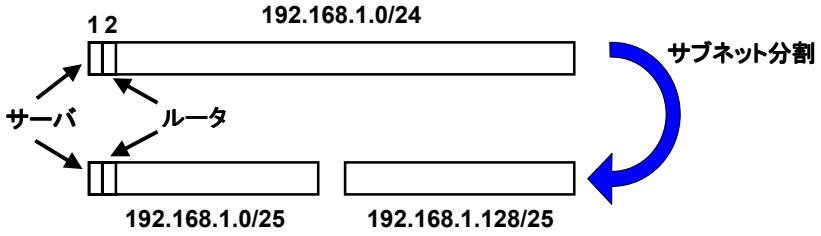
悪い例



- アドレスを先頭と後ろから使用した場合、サブネット分割を行う必要が生じた場合に、リナンバー作業を行う必要が出てしまう。

組織全体でのアドレスの割り当て-2

良い例

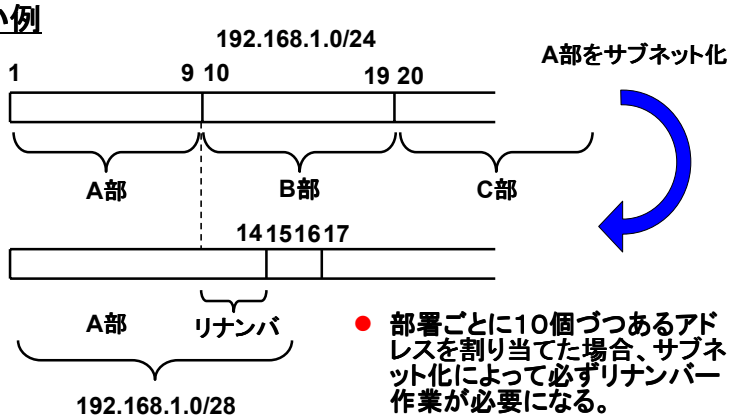


- アドレスを前詰めで使用した場合、サブネット分割を行っても、リナンバー等の無駄な作業を行う必要がない。

IIJ Internet Initiative Japan

部署ごとのアドレスの割り当て-1

悪い例

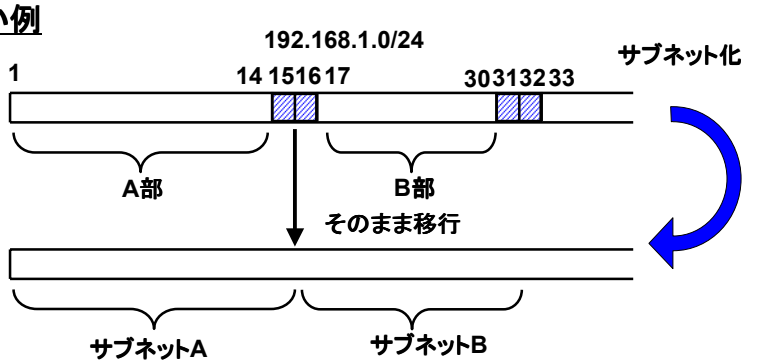


- 部署ごとに10個づつあるアドレスを割り当てた場合、サブネット化によって必ずリナンバー作業が必要になる。

IIJ Internet Initiative Japan

部署ごとのアドレスの割り当て-2

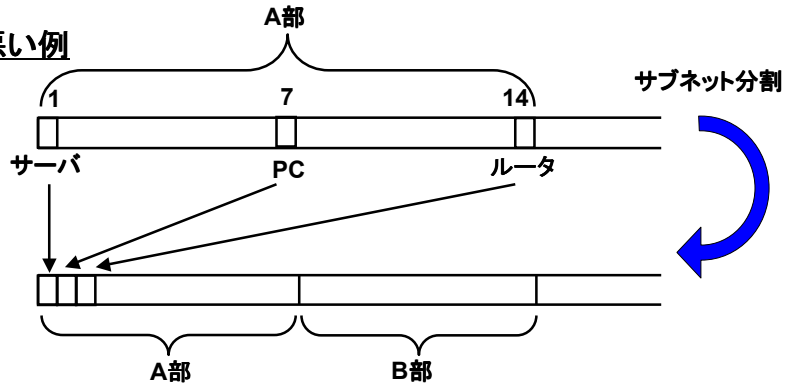
良い例



- 予めサブネットを意識して部署毎にアドレスを割り当てることにより、リナンバー作業を防ぐことができる

部署内でのアドレスの割り当て-1

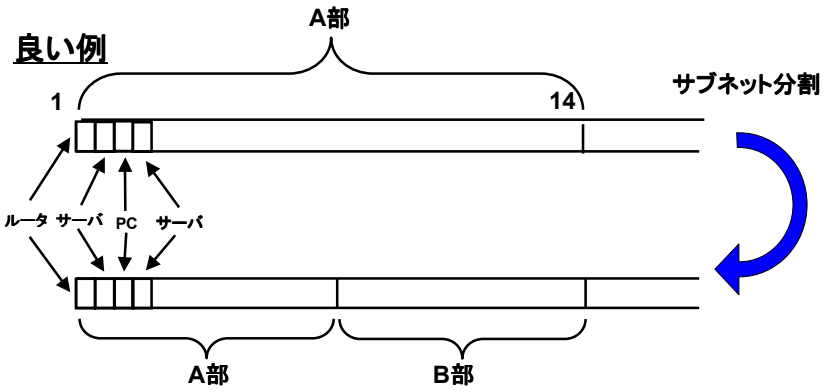
悪い例



- 部署内でルータやサーバ等利用目的別に割り当てるアドレス空間を決めてしまうと、さらなるサブネット化に対応できず、リナンバー作業が発生してしまう。

部署内でのアドレスの割り当て-2

良い例

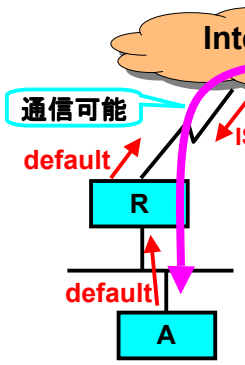


- アドレスを前詰めで使用すればさらなるサブネット化にもスムーズに対応できる

アドレスの割り当てポリシーとは

- アドレスを先頭から詰めて使用するべきか、それとも先頭と後ろから使用していくべきか
  - 先頭から詰めて使用する
- 各部署に割り当てる時は、どのように割り当てていけばいいのか
  - サブネット化を考慮して、例えばA部に1～14、B部に17～30のように割り当てる
- 各部署内で各ホストに割り当てる場合にどのように割り当てていけば良いのか
  - 先頭から前詰めで使用する

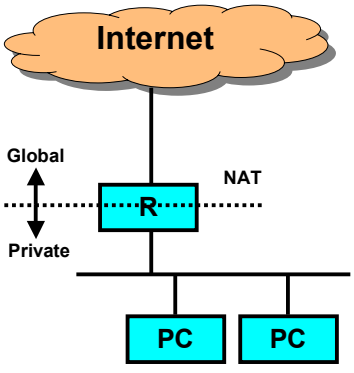
インターネットへの接続形態



- Defaultとは
  - コンピュータ、インターネット業界用語
  - 特に指定が無い場合に選択される指定のこと
  - 債務不履行ではない
- Internet接続したルータに defaultルートを向けることにより、internet上のサーバ群と通信が行える
- Internet接続にはルーティングは必須

IIJ Internet Initiative Japan

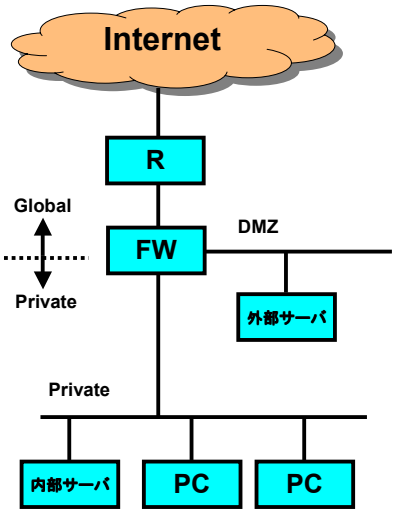
インターネット接続事例—A



- サーバレス運用
  - ISPのDNS,Mail,Webサービスを利用
- セキュリティ
  - 低い
  - ルータのNAT機能に依存
  - 重要なデータをPCに置けない

IIJ Internet Initiative Japan

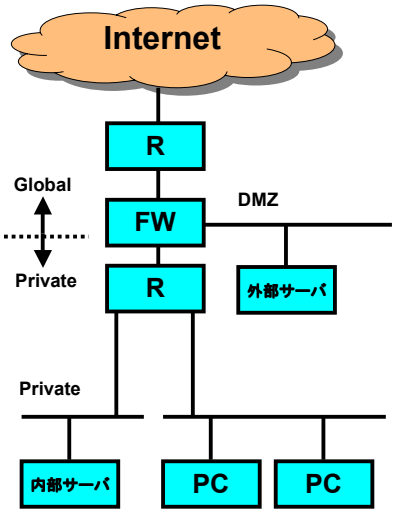
インターネット接続事例ーB



- 自社サーバ運用
- 外部サーバ
  - DMZ(Demilitarized Zone)に設置
  - 公開Web
  - 外部DNS
- 内部サーバ
  - Mail
  - 内部DNS
- セキュリティ
  - 標準的

IIJ Internet Initiative Japan

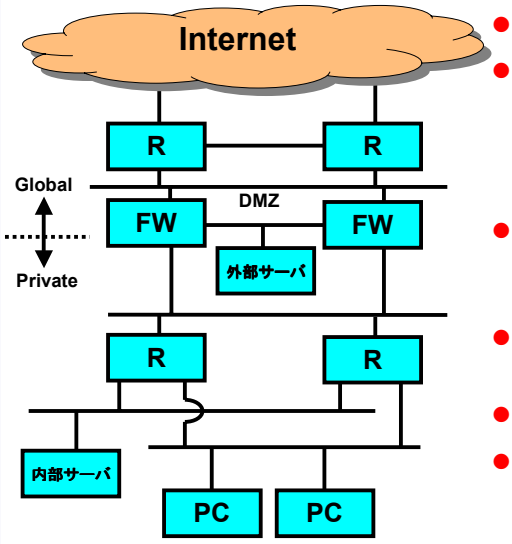
インターネット接続事例ーC



- 自社サーバ運用
- 外部サーバ
  - DMZ(Demilitarized Zone)に設置
  - 公開Web
  - 外部DNS
- 内部サーバ
  - Mail
  - 内部DNS
- セキュリティ
  - 標準的
- 内部サーバの保護

IIJ Internet Initiative Japan

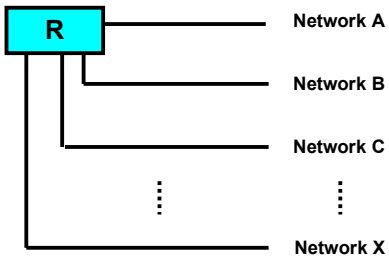
インターネット接続事例ーD

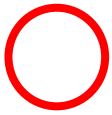


- 自社サーバ運用
- 外部サーバ
  - DMZ(Demilitarized Zone)に設置
  - 公開Web
  - 外部DNS
- 内部サーバ
  - Mail
  - 内部DNS
- セキュリティ
  - 標準的
- 内部サーバの保護
- 回線、機器の二重化
  - HSRP,VRRPの利用

IIJ Internet Initiative Japan

ネットワーク拡張(スター型)



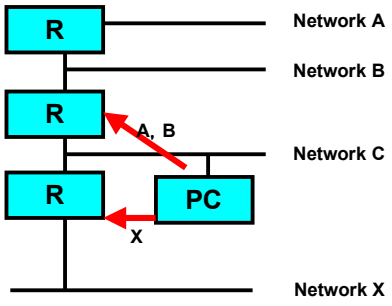


ネットワーク拡張の基本であり、特別な事情がない限り、まずこの形式を検討すべきである

- スター型拡張
  - スター型のネットワーク拡張はルーティングを単純化できるだけでなくポリシー制御も容易なため、小規模から大規模まで幅広く利用されれいる
- 特徴
  - ルーティングが容易
  - ポリシー制御が容易
  - 大規模となると集約されるルータを高性能化する必要がある
  - 多くのネットワークを収容できるルータが必要となるが、VLANなどの利用で安価に構成できるようになった

IIJ Internet Initiative Japan

ネットワーク拡張（数珠型）

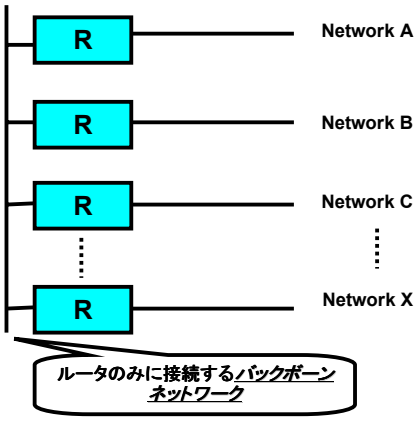


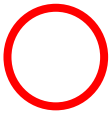
物理的にこの形式しか組めない場合を除いて避けるべき構成である

- 数珠型拡張
  - フロアやビル間などを1つのネットワークで構成し、かつ、そのネットワーク上にクライアントが繋がれるモデル
- 特徴
  - 大規模になるにつれてルーティングが複雑になる
  - ダイナミックルーティングとスタティックルーティングが混在し、誤動作する恐れがある

IIJ Internet Initiative Japan

ネットワーク拡張（L2バックボーン型）



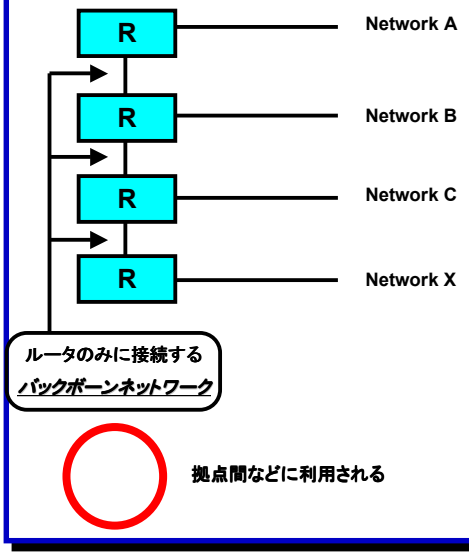


同一構内などのLAN接続などに有効に利用できる

- L2バックボーン型拡張
  - 1つのLayer 2をルータが共有し、PCやサーバとルータを混在させないようにする。
- 特徴
  - ルーティングはルータのみで行えるため、スタティックからダイナミックまで拡張が可能
  - 1つのLayer 2を共有するため、長距離の伝送が難しい
  - 1つのLayer 2が大きくなりすぎる前にバックボーンの階層化を検討する必要がある

IIJ Internet Initiative Japan

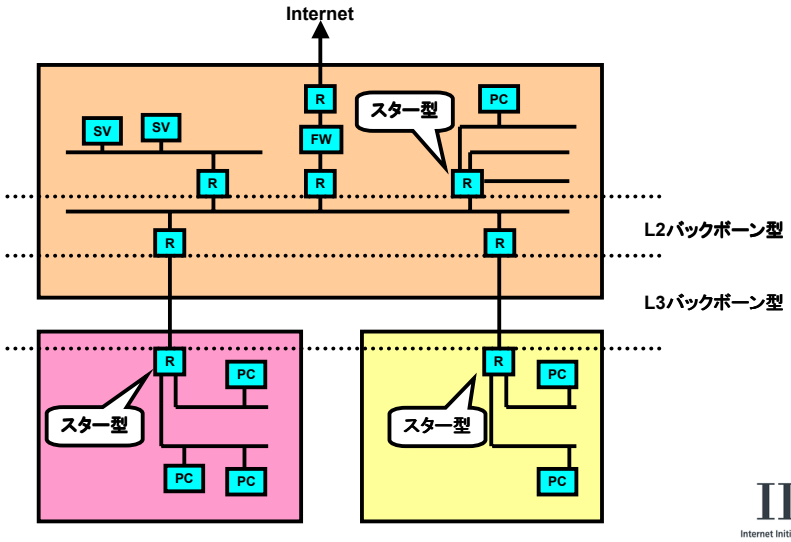
ネットワーク拡張（L3バックボーン型）



- L3バックボーン型拡張
  - ルータ間をPoint to Point ネットワークで接続し、るーたのみのバックボーンネットワークを構築する
- 特徴
  - ルーティングはルータのみで行えるため、スタティックからダイナミックまで拡張が可能
  - 専用線などが利用でき、長距離の伝送が容易
  - L2バックボーンに比べて高価なため、拠点間などの長距離に用いる

IIJ Internet Initiative Japan

ネットワーク拡張事例



IIJ Internet Initiative Japan

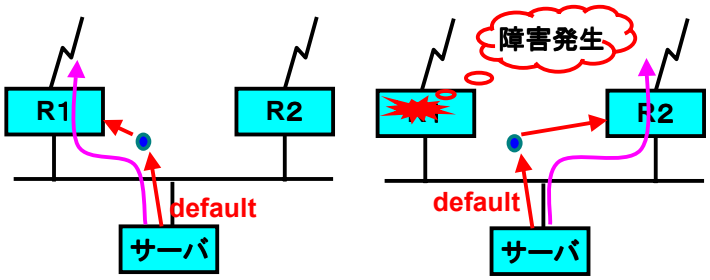
ネットワーク拡張のまとめ

- 小規模な同一構内のネットワークにはスター型を用いる
- 中規模の同一構内のネットワークにはL2バックボーン型を用いる
- 拠点間を結ぶネットワークにはL3バックボーン型を用いる
- 数珠型接続はできる限り避けるようにする

ネットワーク構築に利用される冗長化の仕組み

- STP(spanning tree protocol)
  - レイヤ2での冗長構成
  - 障害の発生から spanning tree変更までには10秒～60秒程度必要
- I/F downと static
  - I/Fの downを検出するとその i/fに向いているroutingが消えることを利用したbackup
  - Ethernet専用線等のdownしないI/Fでは利用できない。
- HSRP/VRRP
  - 一つの仮想的な MACアドレスを複数のルータで共有することで、サーバ等でダイナミックルーティングを利用せずに障害時の切り換えを行う

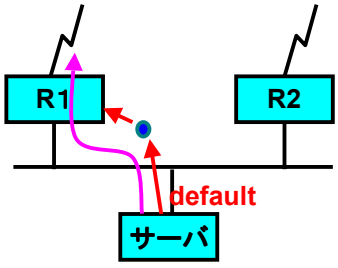
HSRP/VRRP-1



- 障害時には仮想MACアドレスがR1からR2に切り替わる
  - スイッチ等にルータを接続している場合には、ポート、MACアドレスの対応に食い違いが生じるため、さらに切り換えに時間を要する場合がある

IIJ Internet Initiative Japan

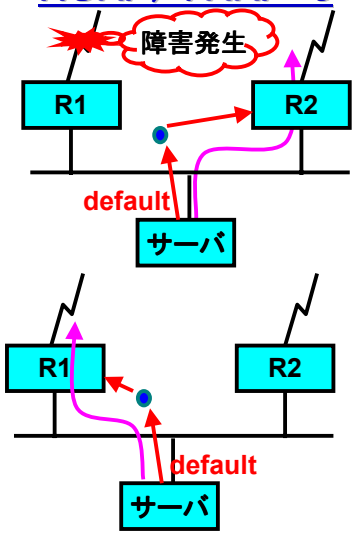
HSRP/VRRP-2



- HSRP+Interface Tracking（通常運用時）
- Interfaceの downを検出して、Trackingすることで回線障害時にactiveルータの切り換えを行う

IIJ Internet Initiative Japan

HSRP/VRRP-3

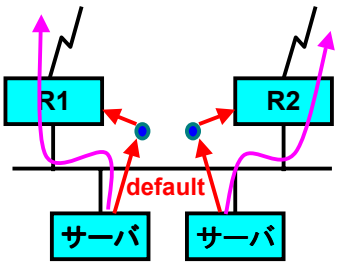


- HSRP+InterfaceTracking（障害発生時）
  - Interface Trackingにより切り替え
- HSRP+Interface Tracking（障害復旧時）
  - 復旧により切り戻しが発生

IIJ Internet Initiative Japan

MHSRP-1

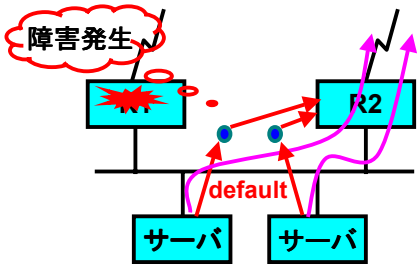
- マルチグループを用いて MHSRPを利用すれば、サーバ毎にトラフィックを分ける事ができる



- MHSRP (通常運用時)
  - それぞれのサーバは対応する HSRPの仮想アドレスにdefaultを向ける

IIJ Internet Initiative Japan

MHSRP-2



- MHSRP (障害発生時)

- なお、MHSRPにはグループID衝突問題があるため、オープンなネットワークでの利用には注意が必要

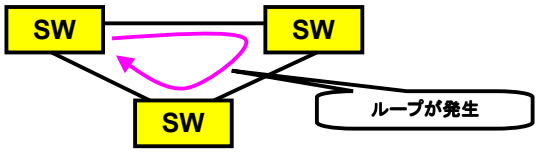
IIJ Internet Initiative Japan

HSRP/VRRPまとめ

- 一つの仮想的な MACアドレスを複数のルータで共有することで、サーバ等でダイナミックルーティングを利用せずに障害時の切り換えを行う
- HSRP
  - Hot Standby Router Protocol
  - RFC2281 (Informational)
  - Cisco社のパテント
- VRRP
  - Virtual Router Redundancy Protocol
  - RFC2338 (Proposed Standard)
  - ルータやファイアウォールなどに実装されている
- MHSRP
  - 1つのネットワークに複数のHSRPを同時利用し、不可分散することが可能

IIJ Internet Initiative Japan

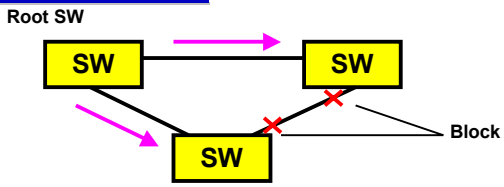
STPの動作-1



- STPを利用しない場合
  - STPを利用せずにスイッチの冗長化するとループが発生する
- ループの発生により様々な問題が発生
  - 各スイッチのアドレステーブルに混乱が生じる
  - 同じフレームが二重に届き、上位層の異常な動作につながる
  - Broadcast floodが発生する

IIJ Internet Initiative Japan

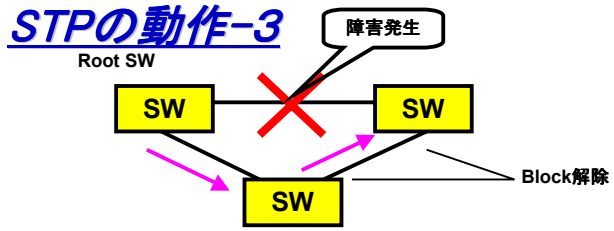
STPの動作-2



- STPを利用する
  - STPの利用によりRoot SWからツリーが構成され、冗長経路はブロッキングされる
  - ブロッキングによりループを防ぐ
- ブロッキングとは
  - 通信が止められている状態
  - ただし、STPのHelloは通信されている

IIJ Internet Initiative Japan

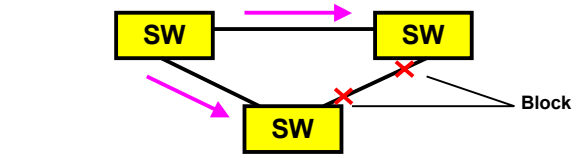
STPの動作-3



- 障害発生が発生した場合
  - 障害発生によりブロッキングが解除され、バックアップされる

IIJ Internet Initiative Japan

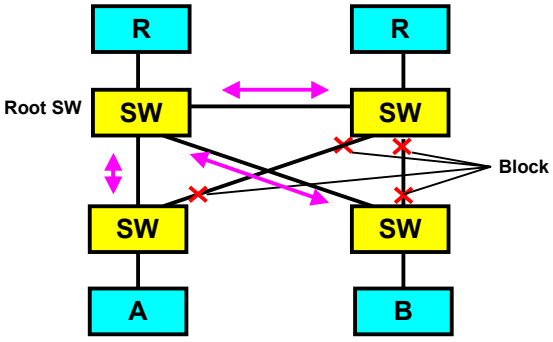
STPの動作-4



- 障害が復旧すると
  - 再びSTPによりRoot SWからツリーが構成され、冗長経路はブロッキングされる

IIJ Internet Initiative Japan

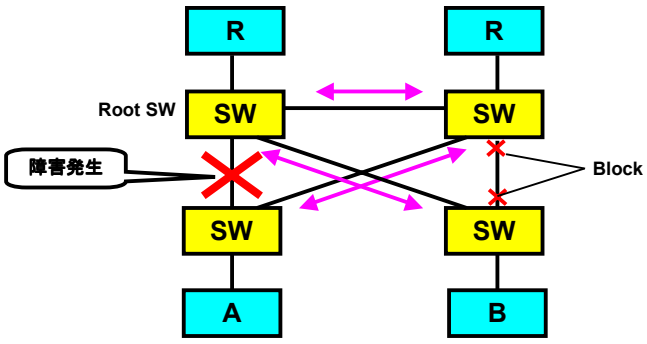
STP事例-1



- 全てのSWをRoot SWに最短となるように設計
- STPにより冗長化経路をブロッキングする

IIJ Internet Initiative Japan

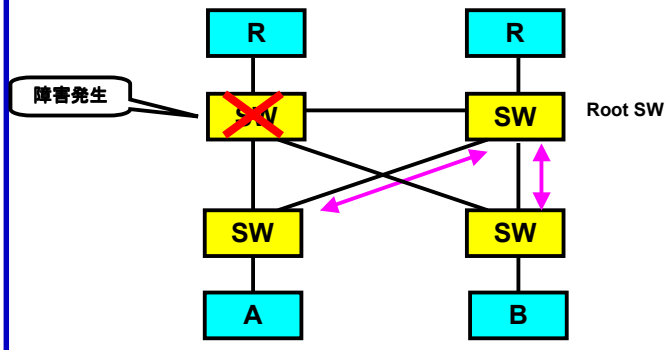
STP事例-2



- 障害が発生するとSTPによりバックアップ経路に切り替わる

IIJ Internet Initiative Japan

STP事例-2



- SWに障害が発生した時もSTPによりバックアップされる
- Root SWに障害が発生した場合には切り替えに時間がかかる

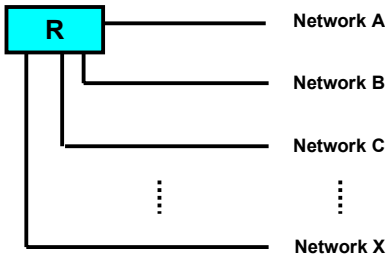
IIJ Internet Initiative Japan

STPまとめ

- STP（Spanning Tree Protocol）
- IEEE 802．1Dの中で定義されている
- データリンク層（L2）プロトコル
- ルートSWからツリーを構成する
- ブロッキングによりループを防止する
- 遠隔地への伝送時などに有効に利用される
- ルートSWはMACなどにより決定されるが、設定することも可能

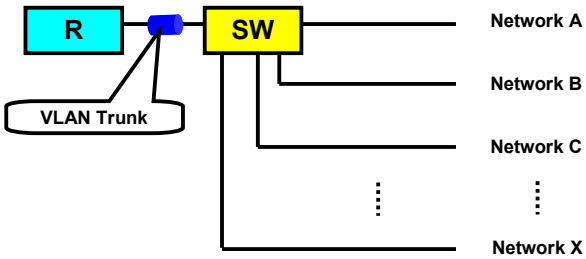
IIJ Internet Initiative Japan

VLAN Trunk-1



- VLAN Trunkしない場合
  - 多くのネットワークを接続する場合、VLAN Trunkを利用しないとルータに多くのインターフェースを用意する必要があり、コストがかかる

VLAN Trunk-2



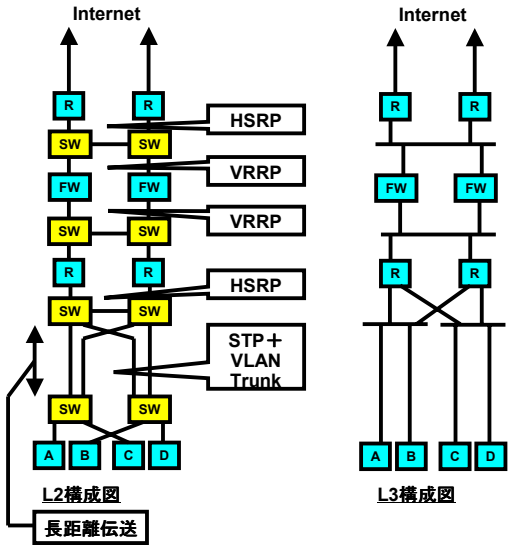
- VLAN Trunkを利用する
- VLAN Trunkによりルータのインターフェースが1つであっても仮想的に複数のインターフェースがあるように見せることができる
- ルータと比較して安価なスイッチのポートをルータのインターフェースとして見せることができる

VLAN Trunkまとめ

- VLAN
  - Virtual LAN
  - 1つのスイッチ内の異なるLANの扱いをVLANと呼ぶ
  - VLAN Trunk、タグVLANのことを略してVLANと呼ぶこともある
- VLAN Trunk
  - 複数のVLANを1つのデータリンク層でまとめて通信する
  - 「タグ付VLAN」「タグVLAN」とも呼ばれる
  - IEEE 802.1Qで定義されている
  - メーカー独自のものも存在する
- VLAN Trunkによりルータのインターフェースが1つであっても仮想的に複数のインターフェースがあるように見せることができる
- ルータに比較して安価なスイッチのポートをルータのインターフェースとして見せることができる
- スイッチをカスケードして利用する場合にはスイッチのダウンを検知できなくなることがあるため注意が必要

IIJ Internet Initiative Japan

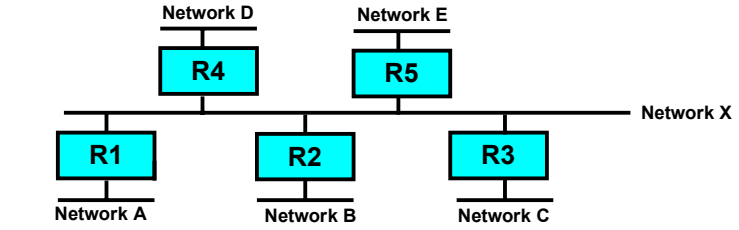
STP+VLAN Trunk事例



- 冗長化されたネットワーク
  - HSRP/VRRP
  - STP
- VLAN Trunkを利用
- 長距離伝送の冗長化とコスト削減を実現
- L3構成図とL2構成図の違いとポイント
  - STPはL2冗長化プロトコルであるため、L3構成図には表れない
  - VLAN Trunkに関してもL同様にL3構成図に表れない

IIJ Internet Initiative Japan

スタティックルーティングの設定



R1

Destination	Next Hop
B	R2
C	R3
D	R4
E	R5

R2

Destination	Next Hop
A	R1
C	R3
D	R4
E	R5

R3

Destination	Next Hop
A	R1
B	R2
D	R4
E	R5

R4

Destination	Next Hop
A	R1
B	R2
C	R3
E	R5

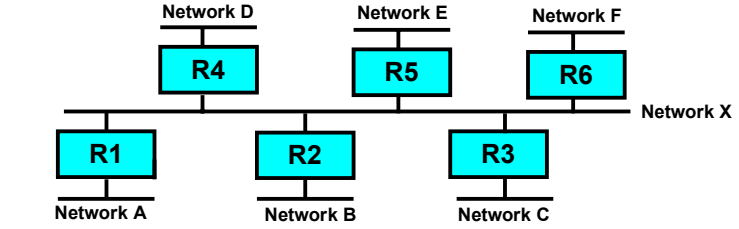
R5

Destination	Next Hop
A	R1
B	R2
C	R3
D	R4

- スタティックルーティングはそれぞれのルータに設定する

IIJ Internet Initiative Japan

スタティックルーティングの追加



R1

Destination	Next Hop
B	R2
C	R3
D	R4
E	R5
F	R6

R2

Destination	Next Hop
A	R1
C	R3
D	R4
E	R5
F	R6

R3

Destination	Next Hop
A	R1
B	R2
D	R4
E	R5
F	R6

R4

Destination	Next Hop
A	R1
B	R2
C	R3
E	R5
F	R6

R5

Destination	Next Hop
A	R1
B	R2
C	R3
D	R4
F	R6

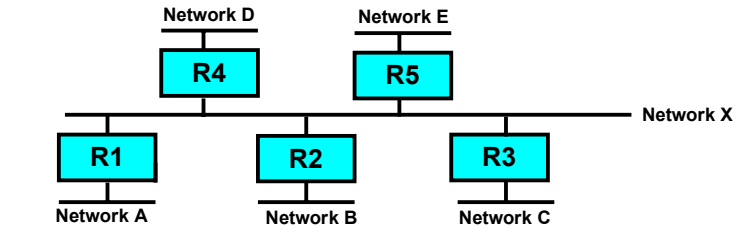
R6

Destination	Next Hop
A	R1
B	R2
C	R3
D	R4
E	R5

- ネットワークが追加されると全てのルータに設定を追加する必要がある

IIJ Internet Initiative Japan

ダイナミックルーティングの設定



R1

Protocol	Net
Ospf	X
Ospf(p)	A

R2

Protocol	Net
Ospf	X
Ospf(p)	B

R3

Protocol	Net
Ospf	X
Ospf(p)	C

R4

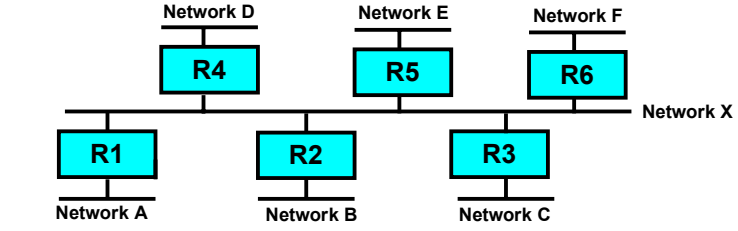
Protocol	Net
Ospf	X
Ospf(p)	D

R5

Protocol	Net
Ospf	X
Ospf(p)	E

- ダイナミックルーティングの設定は利用するプロトコルとネットワークを指定する

ダイナミックルーティングの追加



R1

Protocol	Net
Ospf	X
Ospf(p)	A

R2

Protocol	Net
Ospf	X
Ospf(p)	B

R3

Protocol	Net
Ospf	X
Ospf(p)	C

R4

Protocol	Net
Ospf	X
Ospf(p)	D

R5

Protocol	Net
Ospf	X
Ospf(p)	E

R6

Protocol	Net
Ospf	X
Ospf(p)	F

- ネットワークが追加された場合には追加されたネットワークが接続されているルータのみに設定すればよい

ルーティング設定まとめ

- ルータを導入する際にはルーティング設定が必要
- スタティックルーティングの場合はバックボーンに新しいルータ、ネットワークが接続されると同じバックボーンを利用しているルータ全てに設定を行う必要がある
- ダイナミックルーティングを導入すると新規導入するルータにのみ設定を入れればよい
- 中規模、大規模のネットワークにはダイナミックルーティングを導入したほうが良い

IIJ Internet Initiative Japan

経路制御解説

- ここではダイナミックルーティングの原理について解説します
- 静的経路制御（スタティック）、動的経路制御（ダイナミック）の特徴
- ダイナミックルーティングの動作原理
- ダイナミックルーティングの種類、特徴
- RIP解説
- VLSM
- OSPF解説
- トラブルシューティング

IIJ Internet Initiative Japan

静的な経路制御と動的な経路制御

- 静的（スタティック）な経路制御の特徴
  - 手作業により固定的に経路を設定する
  - 安定している
  - トラフィックや伝送障害の影響を受けない
  - ルーティングプロトコルのためのトラフィックが発生しない
- 動的（ダイナミック）な経路制御の特徴
  - 自動的に経路を設定する
  - ネットワークの変化に対応できる
  - 自動的に最適経路を選択できる
  - 自動的にバックアップ経路を選択できる

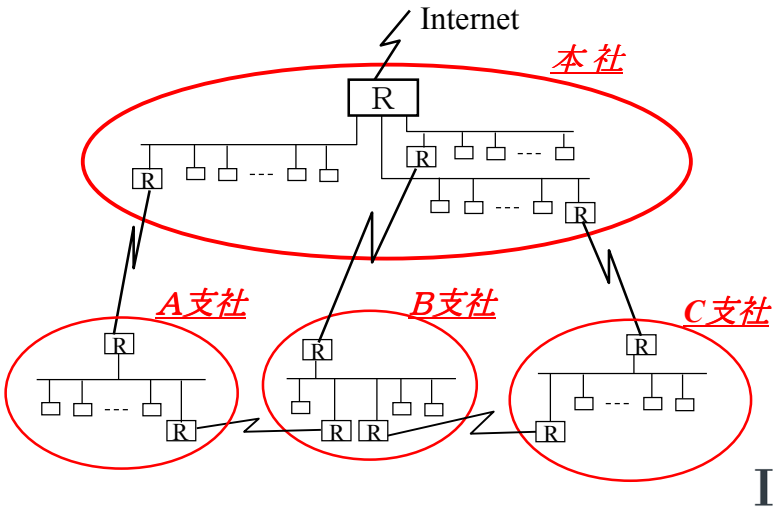
IIJ Internet Initiative Japan

動的な経路制御を選択しなければならない理由-1

- ネットワークの変化に対応しなければならない
  - 一部の追加が全体の変更になることを防ぐ
- 責任者の異なるネットワークを接続する必要がある
  - 複数の管理ネットワークとの接合
- ルータの設定を容易にする
  - 大規模ネットワークを手作業で管理することは難しい

IIJ Internet Initiative Japan

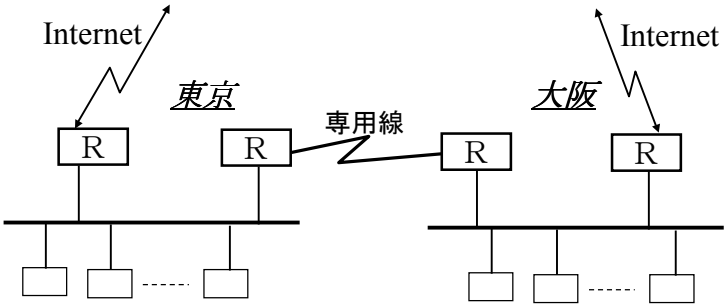
複雑に延びるネットワーク



動的な経路制御を選択しなければならない理由-2

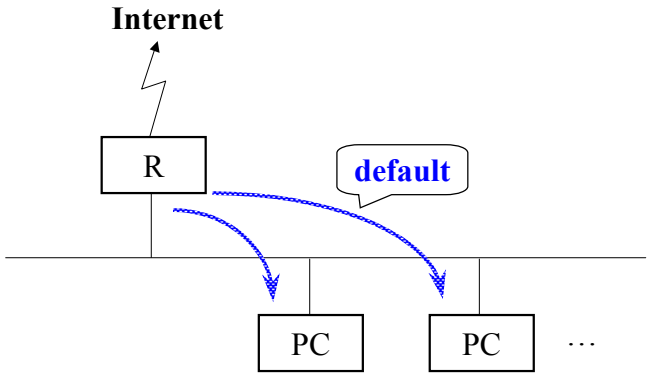
- 自動的に最適経路を選択できる
  - 管理できないほど複雑なネットワークトポロジー
- 自動的にバックアップ経路を選択できる
  - 死守するネットワークが存在する
  - 障害時に強い構成を考える

東京、大阪バックアップ



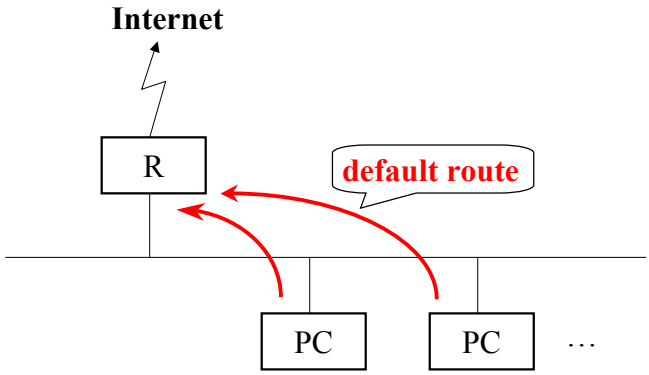
IIJ Internet Initiative Japan

ダイナミックルーティング：経路情報の伝播



IIJ Internet Initiative Japan

ダイナミックルーティング：伝播後の経路情報



ダイナミックルーティングプロトコルの種類

- RIP
  - RFC1058
- RIP2
  - RFC2453
- OSPF
  - RFC2328
- BGP4
  - RFC1771

RIP

- Routing Information Protocol version 1
- RFC1058
- アドレスのみの伝播
  - VLSM使用不可
- ベクトル距離経路制御
- Broadcastのみ
- UNIXに標準添付されている(routed)

RIP2

- Routing Information Protocol version 2
- RFC2453
- netmaskを伝播できる
  - VLSM使用可能
- ベクトル距離経路制御
- RIPと互換性があり、併用も可能
- Multicastを利用可能
  - ホストの軽減を図る
- 最近では対応したroutedがある

OSPF -1

- Open shortest path first
- RFC2328
- Protocol 89
  - TCP(protocol 6)でもUDP(protocol 17)でもない
- netmaskを伝播できる
  - VLSM利用可能

OSPF -2

- Multicast(224.0.0.5/224.0.0.6)を利用する
- Load-balancingを行う
- UNIX標準で添付されていない
  - gated等をインストールする必要がある

BGP4 -1

- Border Gateway Protocol version 4
- RFC1771
- TCP 179
- EGPとしてのEBGPとIGPとしてのIBGPがある
- AS pathの長さにより経路を選択する

BGP4 -2

- 複数の経路が存在する場合は最適経路のみ伝播する
- Load-balancingは行わない
- Updateプロトコルである
- Aggregateできる。Classless Inter-Domain Routing(CIDR)対応

ダイナミックルーティングの解説

- RIPを理解する
  - RIPを理解すれば、OSPF、BGP4を概念的に理解することは容易
- 現場ではいまだにRIPが使用される場合がある
  - OSPFを利用できないルータが存在するため
  - Defaultだけを流すのでRIPで十分
- OSPF解説
  - RIPの知識をベースに解説します

IIJ Internet Initiative Japan

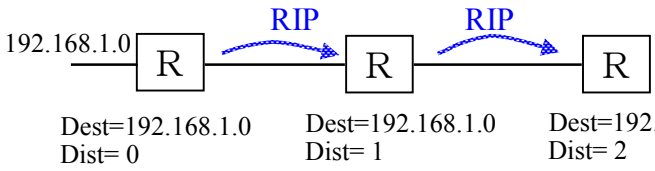
RIPの動作原理 -1

ベクトル距離経路制御
(vector-distance/Bellman-Ford)

vector=destination(ネットワーク)
distance=HOP count(通過したルータの数)

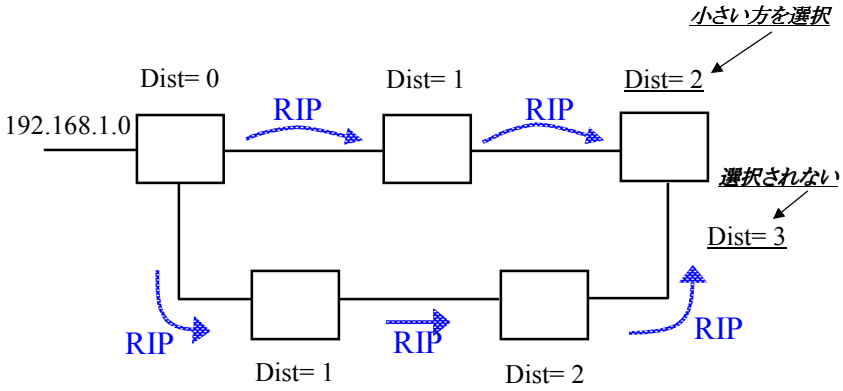
IIJ Internet Initiative Japan

ルータを通る度にdistanceが1追加される



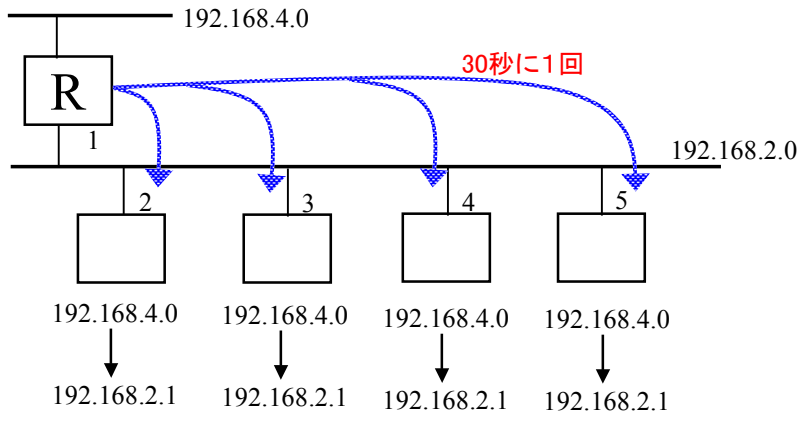
Dest=Destination
Dist= Distance

同じdestinationの場合はdistanceが小さい方を選択



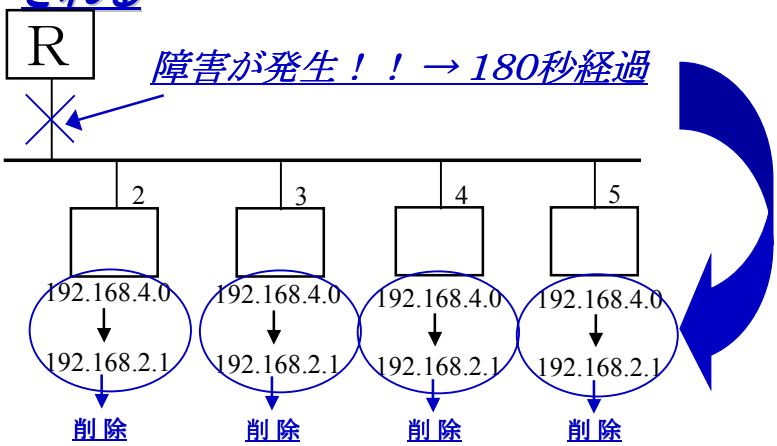
同じDestination同じDistanceの場合は
最初に到着した経路を選択

30秒ごとにbroadcastされる



IIJ Internet Initiative Japan

3分間経路が到着しないと経路は削除される



RIPで得られた経路情報は180秒

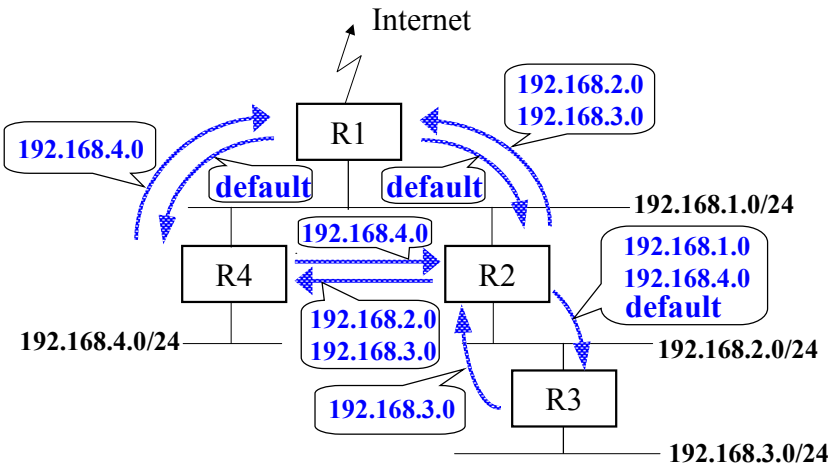
IIJ Internet Initiative Japan

RIPの動作原理-2

- ネットワーク障害時には3分間で経路が切り替わる。 複数ルータがある場合には3分×ルータ数
- RIPはネットマスクを伝播しない
- クラスフルなマスクと見なされる
  - 利用可能な例
    - 192.168.1.0/24
    - 172.16.0.0/16
    - 10.0.0.0/8

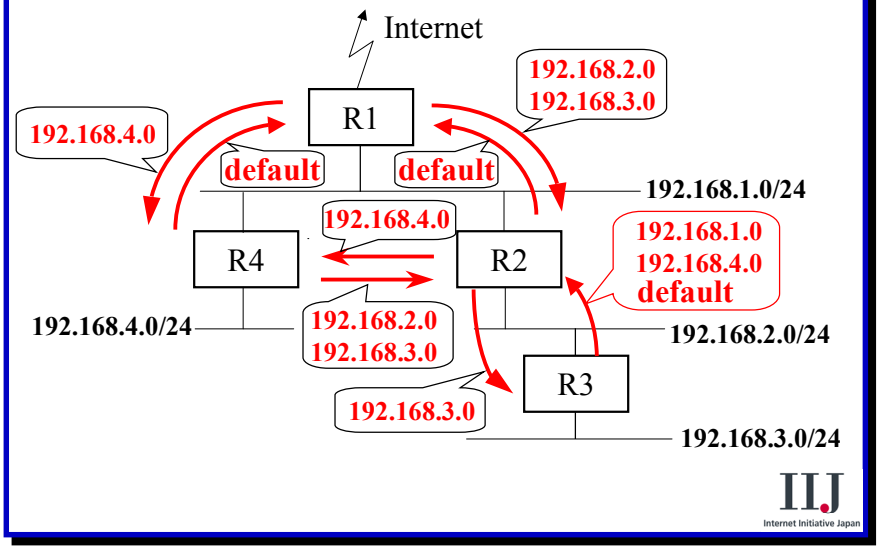
IIJ Internet Initiative Japan

RIP伝播



IIJ Internet Initiative Japan

RIP伝播後の経路情報



RIPの動作原理-3

- 利用不可能な例
  - 192.168.1.0/26
  - 172.16.0.0/24
- 0.0.0.0というアドレスはdefaultとして機能する

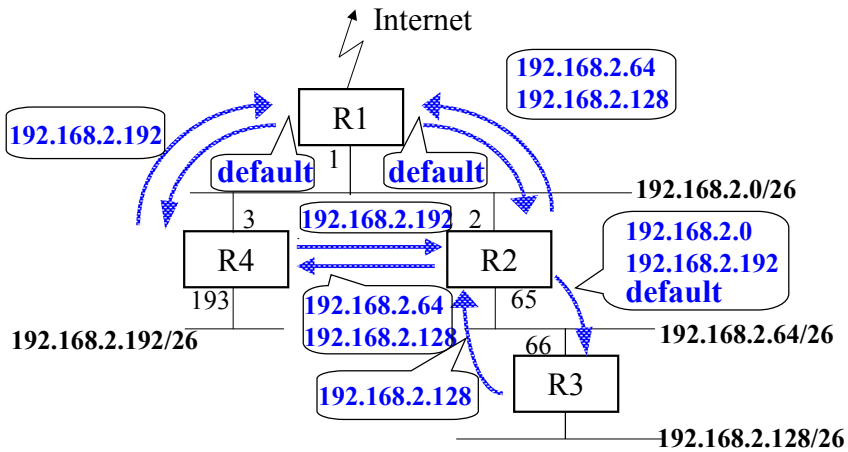
RIPのまとめ-1

- ベクトル距離経路制御(vector-distance/bellman-ford)
  - Vector=destination(ネットワーク)
  - Distance=hop count(通過したルータの数)
- ルータを通る度にdistanceが1追加される
- 同じdestinationの場合はdistanceが小さい方を選択
- 同じdestination同じdistanceの場合は最初に到着した経路を選択

RIPのまとめ-2

- 30秒ごとにbroadcastする
- 3分間経路が到着しないと経路は削除される
- ネットワーク障害時には3分間で経路が切り替わる。
  - 複数ルータがある場合には3分×ルータ数

Subnetmaskありのネットワーク構成



RIPでSubnetmaskを利用する場合-1

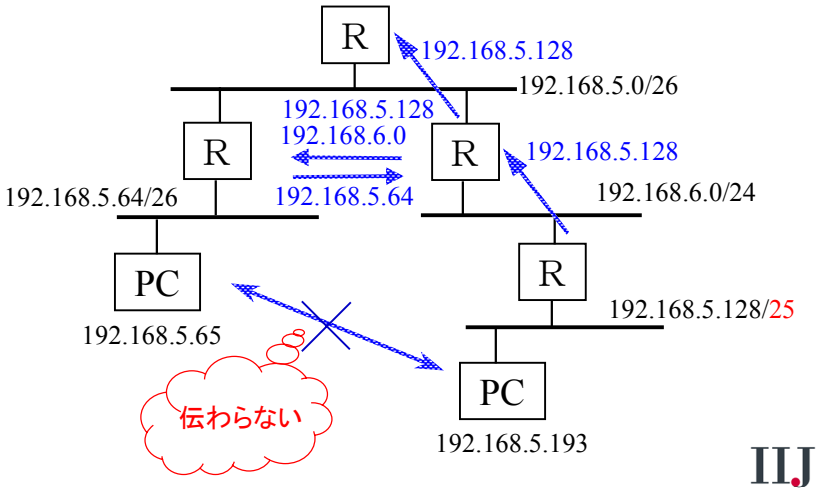
- インターフェースに設定されているnetmaskを適用
- 192.168.2.1/26 ルータのアドレス、マスクの場合

RIPで得られたdestination	ルーティングテーブル
192.168.2.64	192.168.2.64/26
192.168.2.65	192.168.2.65/32
192.168.2.128	192.168.2.128/26
192.168.2.192	192.168.2.192/26
192.168.3.0	192.168.3.0/24
192.168.3.64	192.168.3.64/32

RIPでSubnetmaskを利用する場合-2

- インターフェースに設定されているnetmaskが適用できない場合、RIPでは経路制御できない

VLSMありのネットワーク構成



VLSM(Variable Length Subnet Mask)

- ネットワーク例
  - 192.168.5.0/26
  - 192.168.5.64/26
  - 192.168.5.128/25
- 192.168.5.1が192.168.5.128を受け取った場合
  - 192.168.5.128/26と誤認する
  - 192.168.5.192～192.168.5.255がルーティングされない
- RIPだけではVLSMに対応できない
  - VLSM対応には RIP2、OSPFを利用

IIJ Internet Initiative Japan

ルータでのRIP制御

- 聞く　広告
  - ○　○　RIPのみで運用可能
  - ×　○　defaultのみ広告を行うなどで利用
  - ○　×　defaultを告知しない場合に利用

IIJ Internet Initiative Japan

トラブルシューティング- RIPが伝播しない-1

- 同じbroadcastアドレスを利用していない
  - Broadcastアドレスが異なっている場合
  - 192.168.1.0/24を利用の場合
    - 192.168.1.255　network+all-1
    - 192.168.1.0　network+all-0
    - 255.255.255.255　all-1
    - 0.0.0.0　all-0
- 古いルータやワークステーション等はall-0,all-1固定の場合がある

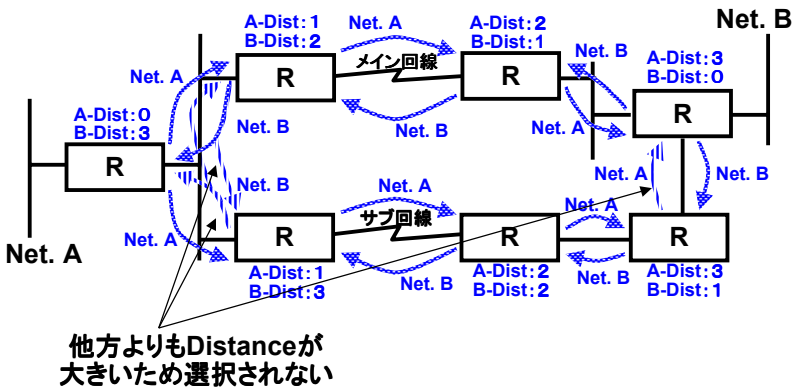
IIJ Internet Initiative Japan

トラブルシューティング- RIPが伝播しない-2

- Broadcastアドレスがfilterされている
  - 255.255.255.255,0.0.0.0などがインターフェースのoutputでfilterされていないか？
- プロトコル、ポートがfilterされている
  - UDP 520がfilterされていないか？
- Unnumberedのi/fでbroadcastを伝播できない
  - unicastで広告するように設定する
  - unicastで広告して良いのか？

IIJ Internet Initiative Japan

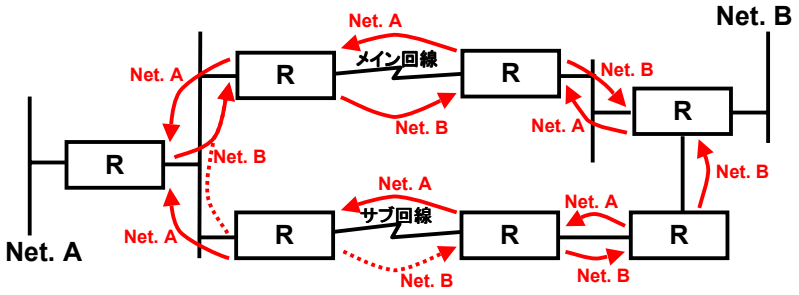
RIPを用いたバックアップ-経路の伝播(定常時)



他方よりもDistanceが
大きいため選択されない

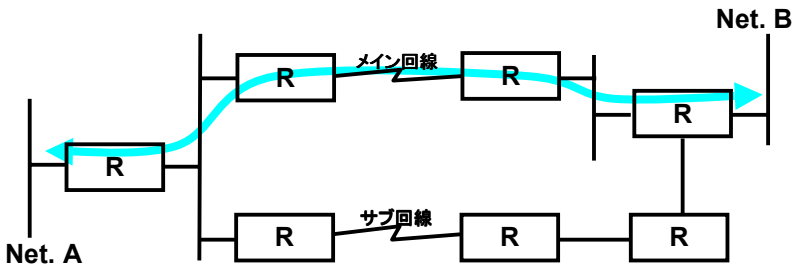
- RIPを利用し、主にバックアップを目的とした構成
- 通常時はメイン回線のみを利用する

RIPを用いたバックアップ -ルーティングテーブル(定常時)



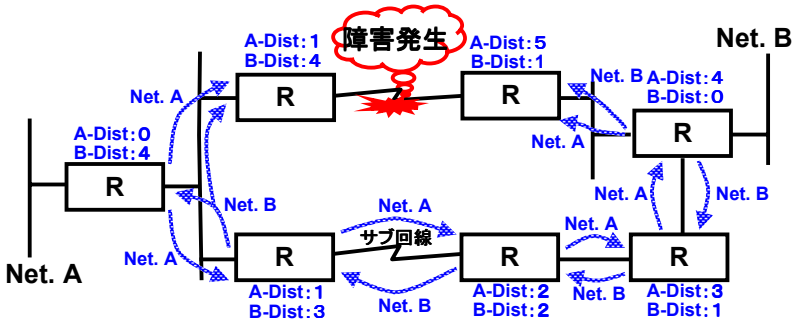
- RIPの経路情報が伝播することにより、各ルータに経路情報が設定される
- Distanceの違いから、メイン回線側の経路が選択される

RIPを用いたバックアップ-トラフィックの流れ（定常時）



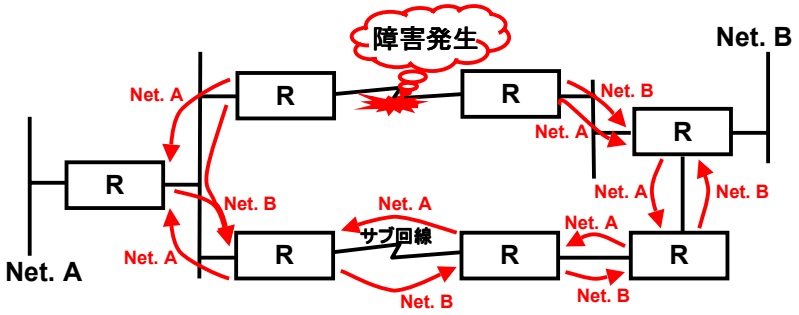
- 通常時はメイン回線のみが利用される

RIPを用いたバックアップ-経路の伝播（障害時）



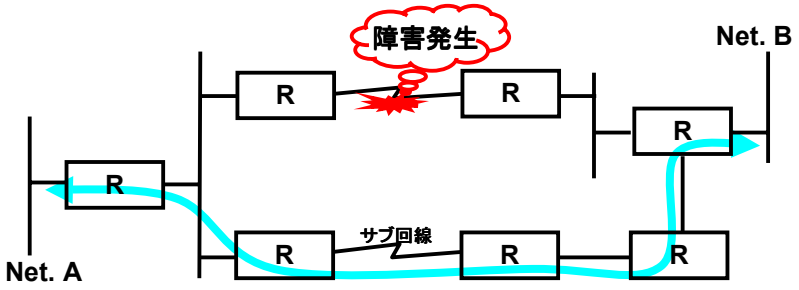
- メイン回線に障害が発生したため、経路情報の伝播が変化する

RIPを用いたバックアップ-ルーティングテーブル(障害時)



- 経路情報の伝播が変化するため、各ルータに設定されている経路情報が変更される

RIPを用いたバックアップ-トラフィックの流れ(障害時)



- メイン回線に障害が発生しているため、トラフィックの流れも変化する
- サブ回線を利用して、通信のバックアップを行う

OSPF解説－1

- 解説方針
  - ここではOSPFを知らない方のために一般的な利用法について解説します。
  - わかりやすさを重視して説明するため、RFCで定義されている厳密なOSPFの定義とは異なる部分もありますが、ご了承願います。
  - 大規模ネットワークではBGPとの連携は欠かせませんが、ここでは説明しません。

IIJ Internet Initiative Japan

OSPF解説－2

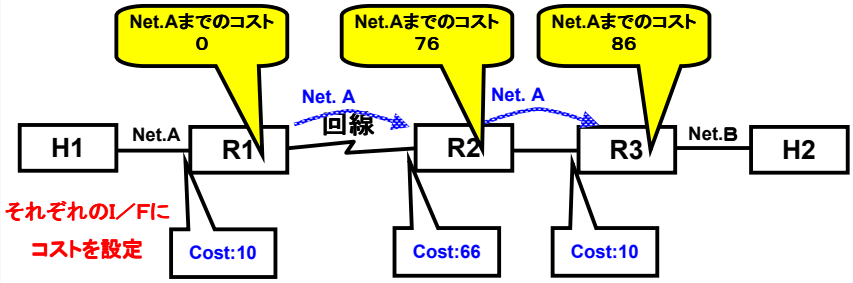
- Link State型ルーティングプロトコル
  - ネットワークトポロジをLSA(Link State Advertisement)と呼ばれる形式でデータベース化し、最適な経路を選択する。
    - RIPやBGPと異なり、単純な経路交換を行なわないため、経路フィルタをかけることは難しい
  - トポロジに変更が合った場合にすぐ変更がかかる
  - ルータ故障検出も可能
    - HELLOパケットによりルータの故障を検出し、バックアップ経路を選択できる。
    - 切り替え時間がRIPよりずっと早い(数秒～1分程度)

IIJ Internet Initiative Japan

OSPFコストとは

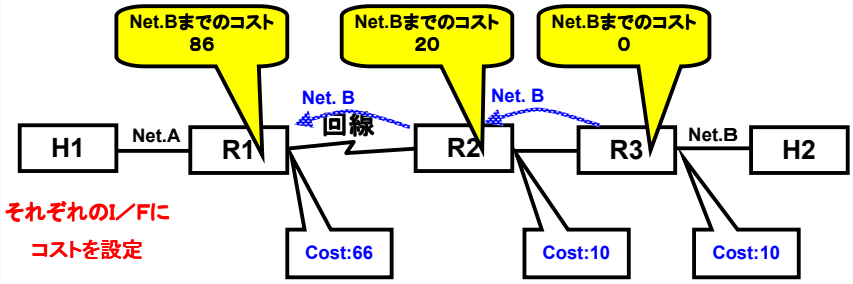
- OSPFではRIPでいうDistanceの変わりにコストを利用する
  - OSPFコストは0～65535の値を取る
  - インターフェース毎に自由にコストを設定することができる
  - コストは小さければ小さいほどネットワーク的に近距離に見せられる
  - ルータによっては回線速度に応じて自動的にコストを付与するものもあるが、ネットワークの高速化などに対応できなくなる恐れがあるため、バックボーンなど重要なインターフェースは明示的に設定したほうが安全

簡単なOSPFコストの計算法－1



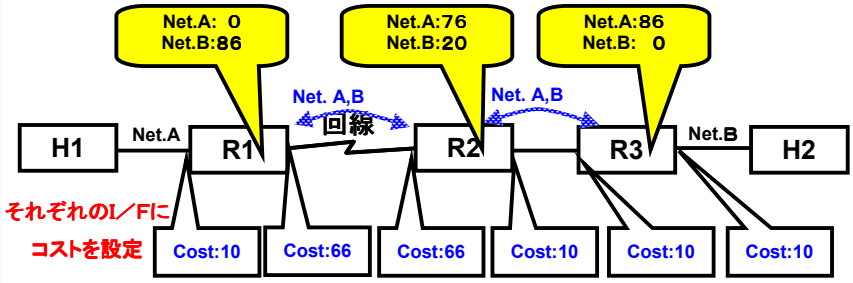
- R1から見たH1への経路
  - R1は直接Net.Aに接続されているため、同じNet.Aに接続されているH1はコスト0として見える
- R2から見たH1への経路
  - R2からは(R1のI/Fに設定されたNet.Aのコスト+R1と接続するI/Fに設定されたコスト)となる
- R3から見たH1への経路
  - R3からは(R2から見たNet.Aのコスト+R2と接続するI/Fに設定されたコスト)となる

簡単なOSPFコストの計算法－2



- R3から見たH2への経路
  - R3は直接Net.Bに接続されているため、同じNet.Bに接続されているH2はコスト0として見える
- R2から見たH2への経路
  - R2からは(R3のI/Fに設定されたNet.Bのコスト+R3と接続するI/Fに設定されたコスト)となる
- R1から見たH2への経路
  - R1からは(R2から見たNet.Bのコスト+R2と接続するI/Fに設定されたコスト)となる

簡単なOSPFコストの計算法－3

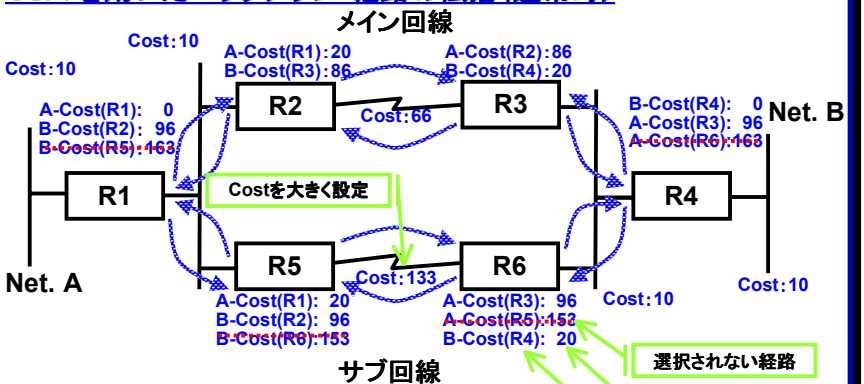


- 同じI/Fに同じコストを付けることにより、行きと帰りのコストを一致させることができる
- 行きと帰りで異なるコストを付与することもできるが、管理が煩雑になるため、理由なく行なうべきではない
- ここで示した図は経路を交換しているように書かれているが、実際はトポロジデータベースの交換により経路を確定している

バックアップ、バランシングを行なうには

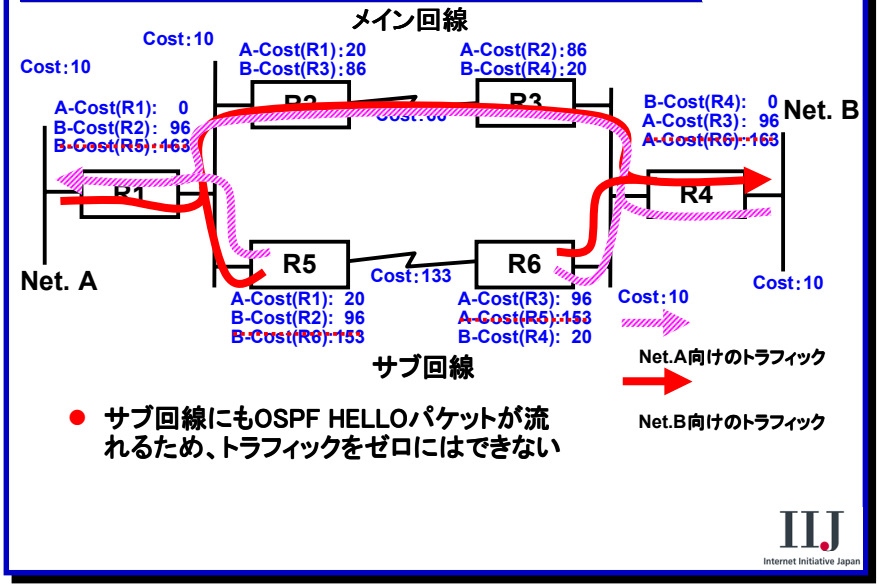
- OSPFでは複数の経路を持った場合にバックアップやバランシングを行なうことができる
- 異なるコストの経路がある場合
  - コストが小さい経路をメインとして利用しコストが大きい経路をバックアップとして利用できる
- 同じコストの経路がある場合
  - バランシングを行ない、トラフィック分散することが可能
  - バランシングを行なっている経路の1つが切断されても残った経路でバックアップすることも可能

OSPFを用いたバックアップ-経路の伝播（通常時）



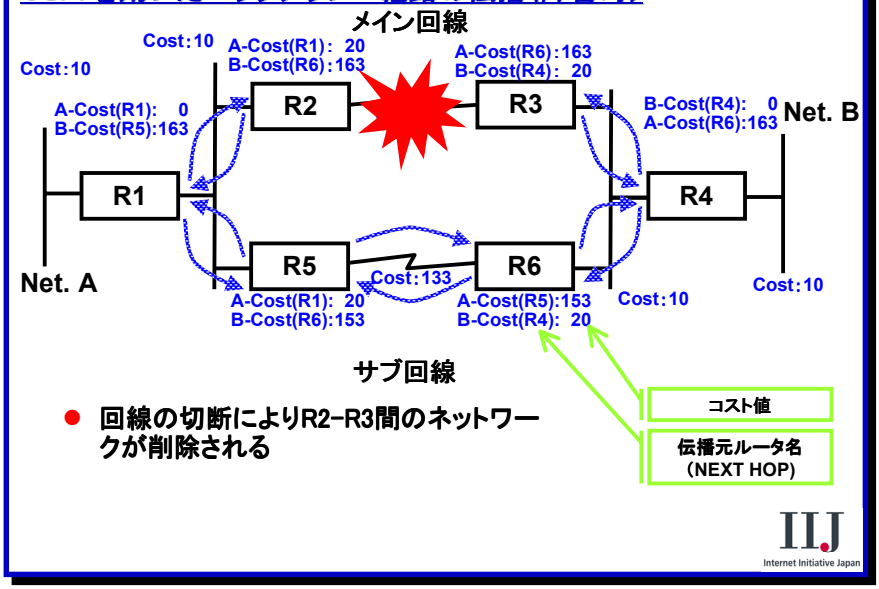
- OSPFを利用して、通常時はメイン回線のみを利用する
- 障害時にはサブ回線を利用してバックアップを行う

OSPFを用いたバックアップ-トラフィックの流れ（通常時）



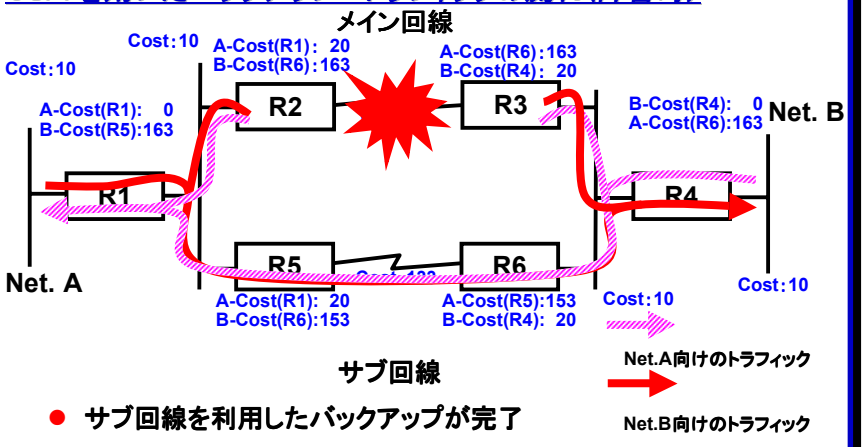
- サブ回線にもOSPF HELLOパケットが流れるため、トラフィックをゼロにはできない

OSPFを用いたバックアップ一経路の伝播（障害時）



- 回線の切断によりR2-R3間のネットワークが削除される

OSPFを用いたバックアップトラフィックの流れ（障害時）

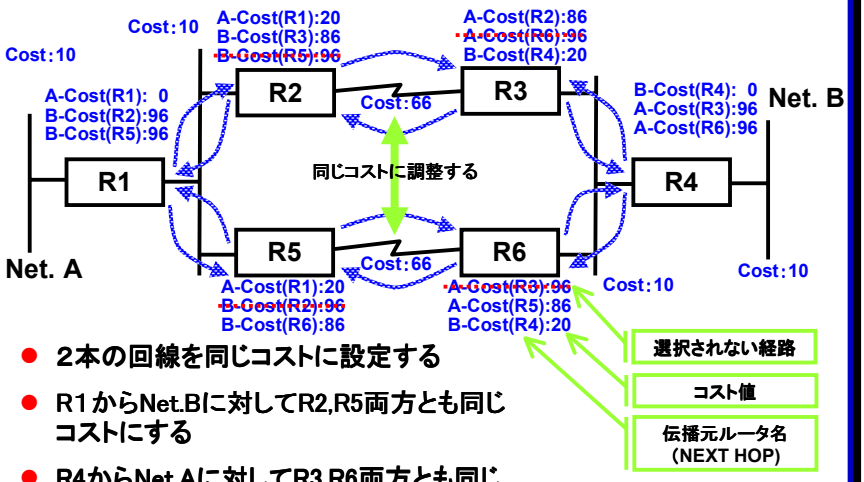


- サブ回線を利用したバックアップが完了

OSPFバックアップルーティングの特徴

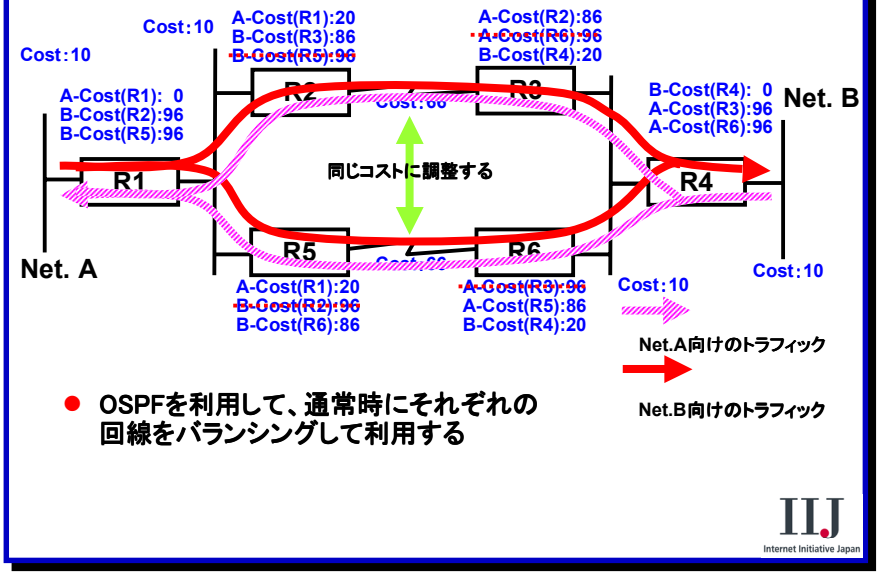
- RIPとは異なり、すばやいバックアップが可能
- バックアップ用の回線上もOSPF HELLOが流れるため、サブ回線を切断することはできない
  - ISDNなどでバックアップさせるにはOSPFだけのチューニングでは難しい
- 2本の回線を別々の用途に利用して障害時にそれぞれバックアップとして利用することが可能

OSPFを用いたバックアップ、バランシング-経路の伝播(通常時)



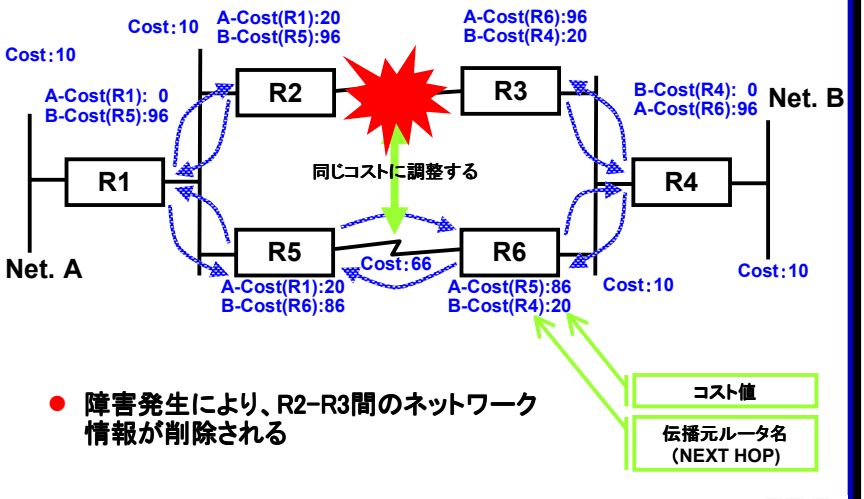
- 2本の回線を同じコストに設定する
- R1からNet.Bに対してR2,R5両方とも同じコストにする
- R4からNet.Aに対してR3,R6両方とも同じコストにする

OSPFを用いたバックアップ、バランシング-トラフィックの流れ(通常時)



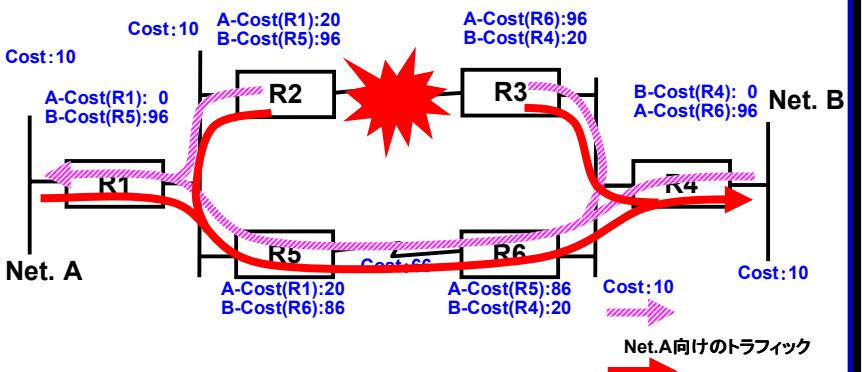
- OSPFを利用して、通常時にそれぞれの回線をバランシングして利用する

OSPFを用いたバックアップ、バランシング-経路の伝播(障害時)



- 障害発生により、R2-R3間のネットワーク情報が削除される

OSPFを用いたバックアップ、バランシング-トラフィックの流れ(障害時)



- 障害が発生していない回線を利用してバックアップを行う

バックアップ、バランシングの特徴

- 障害発生時には50%の帯域でバックアップ
- バランシングは基本的に1:1でバランスするため、速度の異なる回線をバランスさせることは難しい
- 2本の回線を有効に利用し、回線コストを抑えることができる
- LAN等に利用すると100Mbpsメディアを200Mbpsメディアとして利用することもできる

IIJ Internet Initiative Japan

初心者のためのOSPF設定－1

- エリア
  - 必ず0を設定する
    - OSPFでは経路の集約のためにエリアという概念があるが、小規模なネットワークではバックボーンエリア=エリア0だけで構築すればよく、エリアを分けて構築する必要はない
    - エリア0以外のエリアは必ずエリア0と接している必要があるため、むやみにエリア分けをするとバックボーンの拡張が難しくなる
    - ISPなど大規模ネットワークとなるとBGP+OSPFが主流であり、経路の集約という観点ではBGPのほうが優れているため、バックボーン以外のエリアを積極的に使っていくことはあまりない
- デフォルトルート
  - 必ずstaticなどでデフォルトルートを確保してからOSPFでデフォルトルートを流す
    - 余力があればExternal Type 1で流す

IIJ Internet Initiative Japan

初心者のためのOSPF設定－2

- Staticからの経路注入
  - デフォルトルートなどと同じくExternal Type 1で流す
    - OSPFではOSPF以外のstaticやRIPなどから経路を注入するときにExternal Type 1とExternal Type 2が選べるようになっている
    - External Type 1とは
      - 注入時に付与したコストに、注入された場所から実際にOSPFの経路を受けるルータまでのOSPFコストを加えて評価する。同じ経路が複数注入されたときに最も近い出口から出るように制御するために使われる。Staticは注入された個所が最も近いと判断できるため、Type 1が向いている。
    - External Type 2とは
      - 注入時に付与したコストをそのまま維持する。同じ経路が複数注入されたときに注入の際に付けられた優先順位に基づいて評価される。これはBGPなど他のプロトコルの情報をOSPFで実現するために有効な手法だが、現状BGPをそのままOSPFには流せないため、あまり意味がない
      - Ciscoのルータはデフォルト設定がExternal Type 2であるため、注意が必要
    - External Type 1とExternal Type 2を混ぜない
      - OSPFコストとは別にExternal Type 1 > External Type 2という優先順位があるため、障害の切り分けが難しくなる

初心者のためのOSPF設定－3

- ルータID
  - 小規模では特に気にしなくても良いが、loopbackインターフェースを設定したほうが良い。
    - OSPFではルーター間通信にルータID(ルータについているIPアドレス)を用いる。
    - 通常はloopbackインターフェースを設定するとそのアドレスが使われる
    - 同じアドレスを複数のルータのloopbackインターフェースに付けると誤動作するため、注意が必要
- ルータを立ちあげる順番
  - 能力が高く、負荷が低いルータを先に立ち上げたほうがよい。
    - OSPFではDR(Designated Router)「指定ルータ」、BDR(Backup DR)、DROTHERが立ち上がった順に決まり、Ethernetなどマルチアクセスメディアの通信はDRが情報を管理するため、処理能力の余裕があるルータに行なわせたほうが良い。
    - 小規模では意識しなくても問題が発生しないことがほとんど。

トラブルシューティング-RIPv2とOSPFが伝播しない

- ルータのfilter等でmulticastアドレスや、protocol、portなどが制限されていないか注意する
  - RIP2
    - 224.0.0.9
    - UDP 520
  - OSPF
    - 224.0.0.5/224.0.0.6
    - Protocol 89
- Multicastをサポートしない場合
  - OSによってはmulticastを受けられない場合がある このときはbroadcastにて代用する

IIJ Internet Initiative Japan

ダイナミックルーティングのまとめ

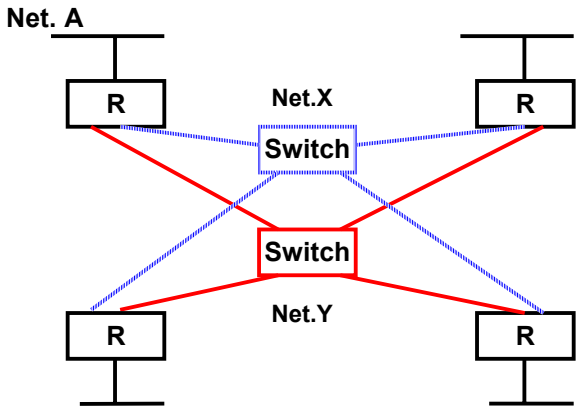
- VLSMを考慮するとRIP2,OSPFへの移行が望まれる
- 単純なネットワーク構成はstaticを選択
- Defaultのみを利用する場合はRIPでも十分
- バランシングなどを行なう場合はOSPFを用いる

IIJ Internet Initiative Japan

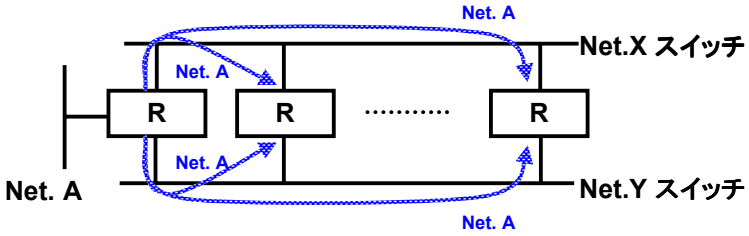
ダイナミックルーティングプロトコルを用いた障害に強いネットワーク構成

- デュアル構成＋OSPFによるバックアップ、バランシング
- リングトポロジによるバックアップ

デュアル構成＋OSPFを用いたバックアップ、バランシング接続図

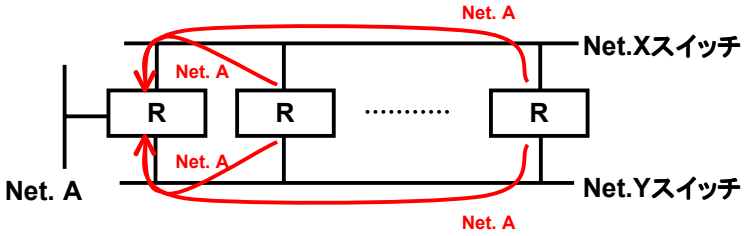


デュアル構成＋OSPFを用いたバックアップ、バランシング-経路の伝播（通常時）



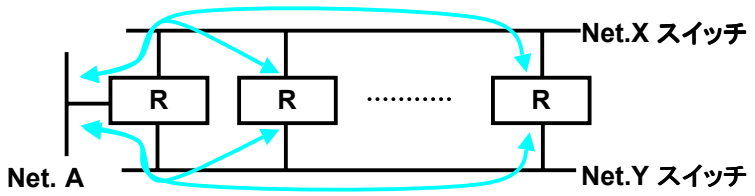
- OSPFで Net.Aの経路情報を広告する
- 経路情報は各ルータに対して、2つのスイッチから等価に伝播する

デュアル構成＋OSPFを用いたバックアップ、バランシング-ルーティングテーブル（通常時）



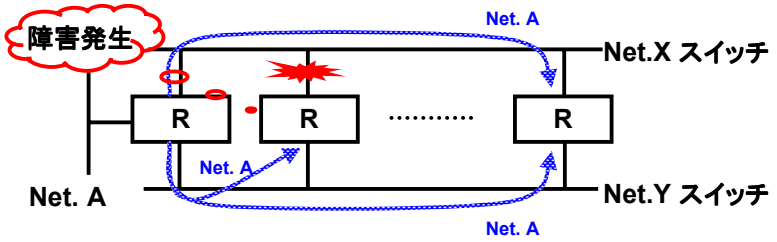
- 伝播した経路情報により、各ルータに経路情報が設定される。
- 2つのスイッチから等価な経路情報が伝播してきたため、2つの経路情報が設定される

デュアル構成＋OSPFを用いたバックアップ、バランシング -トラフィックの流れ（通常時）



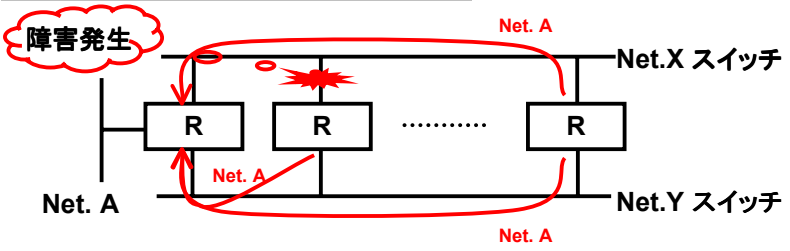
- 通常時には、2つのスイッチを経由するトラフィックがバランスする

デュアル構成＋OSPFを用いたバックアップ、バランシング -経路の伝播（障害時）



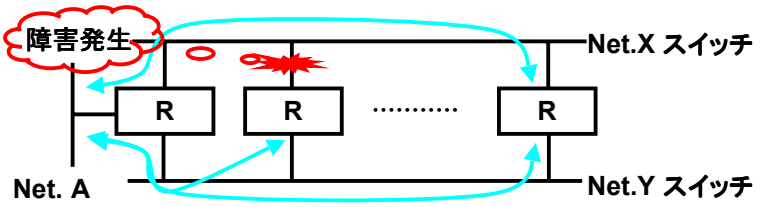
- 障害発生により、経路情報の伝播に一部に変化が生じる

デュアル構成＋OSPFを用いたバックアップ、バランシング -ルーティングテーブル（障害時）



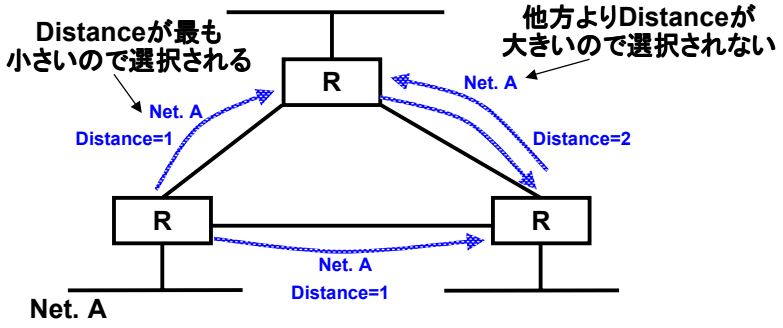
- 伝播する経路情報が変化するため、各ルータに設定されている経路情報も変化する
- 一方のスイッチからの経路が消えても、もう一方のスイッチからの経路でバックアップを行う

デュアル構成＋OSPFを用いたバックアップ、バランシング -トラフィックの流れ（障害時）



- 障害時には、2つのスイッチどちらかを利用して障害を迂回することができる

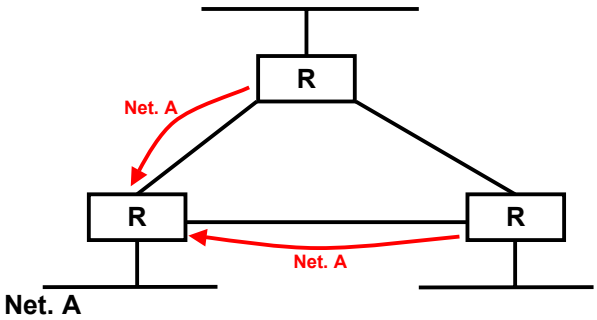
リングトポロジによるバックアップ -経路の伝播(通常時)



- RIPで Net.Aの経路情報を広告する
- 通常時は最短な経路が優先される

IIJ Internet Initiative Japan

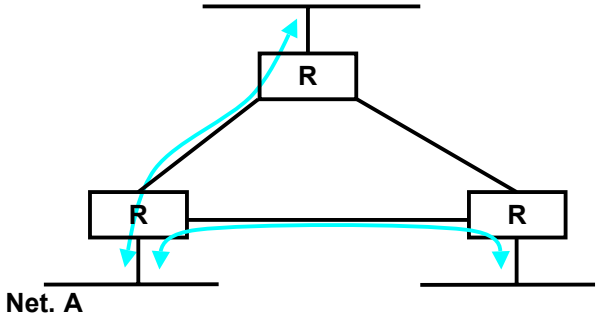
リングトポロジによるバックアップ -ルーティングテーブル(通常時)



- 伝播した経路情報から、各ルータに経路情報が設定される

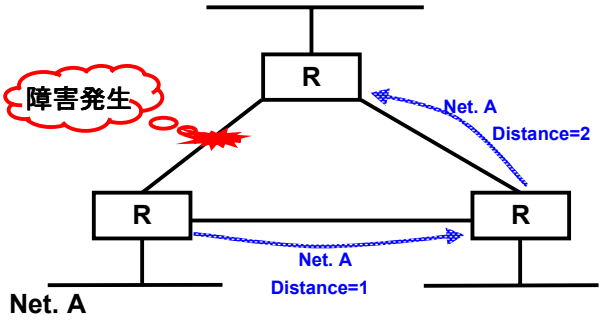
IIJ Internet Initiative Japan

リングトポロジによるバックアップ －トラフィックの流れ（通常時）



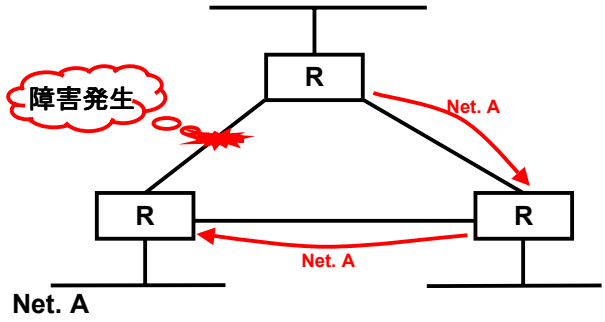
- 通常時は最短な経路が優先されて、通信が行われる

リングトポロジによるバックアップ －経路の伝播（障害時）



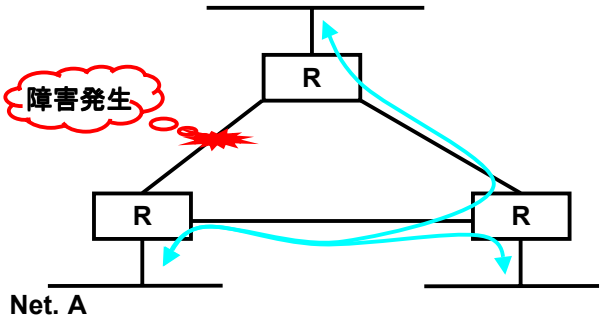
- 障害により、経路情報の伝播に変化が生じる

リングトポロジによるバックアップ -ルーティングテーブル(障害時)



- 伝播する経路情報の変化により、ルータに設定されている経路情報も変化する

リングトポロジによるバックアップ -トラフィックの流れ(障害時)



- 障害時には、遠回りな経路を利用して通信をバックアップする

まとめ－1

- データリンク層とネットワーク層の違い
  - データリンクフレームは中継が起こる毎に変化する
  - IPデータグラムは変化しない
  - データリンクフレームの宛先＝IPデータグラムの宛先とは限らない
- ハブとスイッチ、スイッチとルータの違い
  - それぞれを有効に配置する
- インターネット接続にはルーティングは必須
- サーバなどの安全性を要求されるものは別のセグメントに配置する
- ネットワークの拡張を考慮したアドレス割り当てポリシーで運用する

IIJ Internet Initiative Japan

まとめ－2

- 冗長化のためにSTP、HSRP/VRRPなどを利用する
- VLAN Trunkによりポート単価を下げることが可能
- スタティックルーティングはバックボーンの拡張に伴いダイナミックルーティングに移行する必要がある
- ダイナミックルーティングは基本を理解すれば応用できる
- ダイナミックルーティングを利用すれば障害に強いネットワークを構築できる
- OSPFを利用すればバランシングとバックアップを同時に実現可能

IIJ Internet Initiative Japan