

OSPF と RIP

WIDE

加藤 朗

東京大学大型計算機センター

kato@wide.ad.jp

Agenda

Dec 15, 1998
WIDE Project
Foil 2

WIDE

- 経路制御概論
- RIP, RIP2
- OSPF
- OSPF の運用
- まとめ

概論

IP の経路制御の基礎 (1)

Dec 15, 1998
WIDE Project
Foil 4

WIDE

- 宛先アドレスによる経路選択
 - ルータには経路表を予め設定
 - 宛先アドレスで経路表を検索
 - 出力インターフェースと次段ルータ
- 経路選択の粒度
 - パケット毎に行う
 - 同じ宛先でも別の経路の可能性
 - ルータ単位の経路制御
 - 他のルータとは独立に経路判断
 - 各ルータの経路表の管理が問題
 - hop-by-hop な経路制御

IP の経路制御の基礎 (2)

Dec 15, 1998
WIDE Project
Foil 5

WIDE

- 宛先アドレスのみによる制御
 - 他のフィールドは参照しない
 - 行き
 - 宛先アドレスで経路制御
 - 帰り
 - 宛先アドレスで経路制御
 - 行きと帰りは一般には異なる経路

IP Header

0	1	2	3
01234567890123456789012345678901			
Ver	IHL	TOS	Total Length
Identification		F	Fragment Offset
TTL	Protocol	Header Checksum	
Source Address			
Destination Address			
Options if any			Padding

古典的な IP の経路制御

- 経路の種類
 - ホスト経路
 - ネットワーク経路
 - default 経路
- 経路検索
 - 宛先アドレスの 32bit 全部
 - ホスト経路を検索
 - 宛先のネットワーク部を抽出
 - ネットワーク経路を検索
 - (もしあれば)
 - default 経路を採用

サブネットの導入

Dec 15, 1998
WIDE Project
Foil 8

WIDE

- RFC950 で規定
 - 1985 年 8 月
- Class B アドレスを効率的に使用
 - 「共通な」 subnet mask
 - 可変長 subnet mask は不可
- 経路検索
 - ホスト経路を優先
 - 宛先と同じのネットワーク部ならば
 - Subnet 経路の検索

現在の経路制御

Dec 15, 1998
WIDE Project
Foil 9

WIDE

- CIDR
 - Classless InterDomain Routing
 - RFC1517 で定義 : 1993 年 9 月
 - 経路には「必ず」mask 長が付随
 - implicit な mask 長は廃止
 - Class A/B/C の区別の終了
 - 不連続な netmask も廃止
- 経路の表現
 - 宛先 / mask 長
 - 203.178.136.0/24 : Class C 一つ
 - 203.178.136.0/25 : 前半のみ
 - 203.178.136.0/23 : Class C 二つ

CIDR の背景

Dec 15, 1998
WIDE Project
Foil 10

WIDE

- Class B の推奨
 - 1988 年ぐらいまで
 - 経路一つに 2byte で OK
 - わが国では 133.1 – 133.254 など
- その後
 - インターネットの発達
 - Class B 空間の枯渇の恐れ
 - 複数の Class C の割り当て
 - 経路数の急増

CIDR の狙い

Dec 15, 1998
WIDE Project
Foil 11

WIDE

- 経路経路数の増加への対処
 - 連続する経路を集約
 - 131.112.0.0/16 + 131.113.0.0/16
 - 131.112.0.0/15 一つで表現
 - 経路数の低減
 - 経路の増加の抑制
- プロトコルの改良
 - OSPF → OSPF2
 - RIP → RIP2
 - BGP3 → BGP4

CIDR と経路検索

Dec 15, 1998
WIDE Project
Foil 12

WIDE

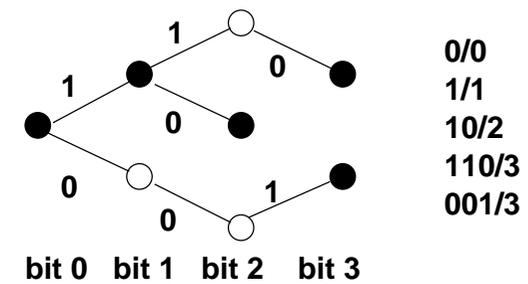
- 宛先アドレスによる経路検索
 - 経路には mask 長が付随
 - 経路の候補の選択
 - $rt \ \& \ mask == dst \ \& \ mask$
 - 最長一致経路を採用
 - 最も mask 長の長いもの
 - longest match / best match

CIDR と経路制御 (2)

Dec 15, 1998
WIDE Project
Foil 13

WIDE

- 二分木による longest match



CIDR と経路検索 (3)

Dec 15, 1998
WIDE Project
Foil 14

WIDE

- 経路表の構成
 - 縮退した二分木
 - Radix Tree
 - 4.3BSD Reno
 - 複雑で backtrack
 - 非連続 netmask に対応
 - TRIE
 - アルゴリズムは簡単

CIDR の効果

Dec 15, 1998
WIDE Project
Foil 15

WIDE

- Global Internet
 - 経路数が少し減少
 - 各種の促進策
 - APNIC の No question asked policy
 - /19 より長い経路は受理しない
 - ・ Sprint Net で実施
 - ・ 最近は /20 か (ARIN 割り当てポリシー)
 - 経路数の増加の抑制
 - ほぼ時間に対して線形増加
 - 最近ではあまり増加せず
 - 現在、約 54,000 経路

CIDR の効果 (2)

Dec 15, 1998
WIDE Project
Foil 16

WIDE

- Enterprise 内部では
 - 可変長 subnet mask (VLSM) の実現
 - IP アドレス空間の有効利用
 - ホスト数に対応した mask 長の設定
- VLSM の実現には
 - CIDR 対応の経路制御方式
 - CIDR 対応の経路制御プロトコル
 - CIDR 対応の管理者

経路の粒度

Dec 15, 1998
WIDE Project
Foil 17

WIDE

- 古典的な IP 経路制御
 - ホスト経路
 - /32
 - ネットワーク経路
 - /8, /16, /24
 - default 経路
 - 0/0

経路の集約 (1)

Dec 15, 1998
WIDE Project
Foil 18

WIDE

- 古典的ネットワーク
 - ネットワーク単位
 - /8 or /16 or /24
- RFC950
 - サブネット単位
 - 例 : /23
 - ネットワーク単位
 - 例 : /16
- CIDR
 - 任意の bit で集約可能

経路の集約の場所

Dec 15, 1998
WIDE Project
Foil 19

WIDE

- ISP の出口
 - インターネットでは必須
 - 割り当てられた CIDR Block 単位
- Enterprise network では
 - 特に必要ない
- 例外
 - 巨大ネットワーク
 - 1000 を越える subnet
 - 集約した方がいい
 - RIP-2
 - 集約は必須

アドレス枯渇

Dec 15, 1998
WIDE Project
Foil 20

WIDE

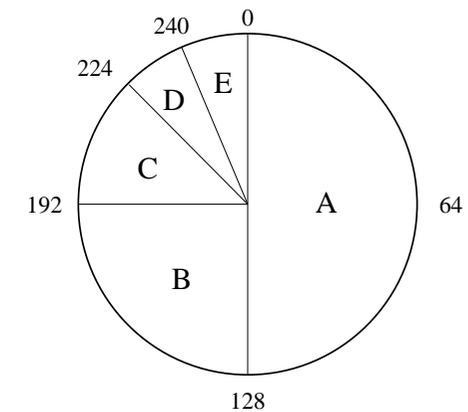
- 控え目なアドレス割り当て
 - 必要最低限の割り当て
 - レジストリとの攻防
- Private アドレス
 - 10.0.0.0/8
 - 172.16.0.0/12
 - 192.168.0.0/16

アドレス枯渇 (2)

Dec 15, 1998
WIDE Project
Foil 21

WIDE

- Class A/B の返還要請
 - Class A 空間の分割割り当て
 - Class C の 4 倍 !
 - 試験的割り当ては既に開始
 - 究極の解決策は IPv6



- IP の経路制御
 - 宛先アドレスによる経路検索
 - 経路表の管理が命
- 静的な経路制御
 - 管理者の設定した経路
 - 安定
 - トポロジ変化には追従せず
- 動的な経路制御
 - 経路制御プロトコル
 - いろいろなオーバーヘッド
 - トポロジ変化に追従
 - インターネットの基本

- Distance Vector (Bellman-Ford) 型
 - 宛先に対する距離 : metric
 - metric の小さい経路を選択
 - ルータが互いの経路表を交換
 - 繰り返して収束
 - count-to-infinity 問題
 - 経路情報とトラフィックは逆方向
 - 選択的な経路送信も可能
 - ポリシの実現

- Link State 型

- トポロジデータベースを作成
 - 仮想的には一つ
 - 実際には各ルータにコピー
 - コピーの同期が問題
- 経路の計算
 - 自ルータを根とする Spanning Tree
 - Dijkstra のアルゴリズム (SPF)
- 収束とループの解消が高速
- フィルタの設定はできない

RIP, RIP2

RIP の概要

Dec 15, 1998
WIDE Project
Foil 26

WIDE

- RFC1058 で定義
 - 「Historic」プロトコル
 - see RFC2400
 - UDP port 520 を使用
 - IP broadcast を使用
- 実は結構広く使われている
 - 古典的なプロトコル
 - BSD Unix の routed

RIP の概要 (2)

Dec 15, 1998
WIDE Project
Foil 27

WIDE

- Distance Vector 型の経路制御プロトコル
 - ルータ相互間での経路情報の交換
- Metric で経路の質を表現
 - 小さい metric が良い経路
 - 1 ~ 16 の範囲
 - 16 は無限大
 - 届かないことの表明
- 宛先 0.0.0.0
 - default 経路
 - 全ての宛先にマッチ

RIP の概念

Dec 15, 1998
WIDE Project
Foil 28

WIDE

- ルータが自分の経路表を順次教える
 - 30 秒に 1 回 flooding
 - ブロードキャスト
 - 180 秒更新されない経路
 - down していると解釈
 - metric を 16 にする
 - さらに 120 秒経つと消去

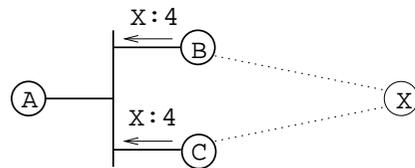
RIP Metric

Dec 15, 1998
WIDE Project
Foil 29

WIDE

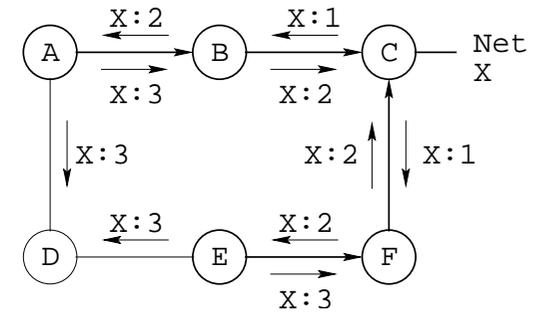
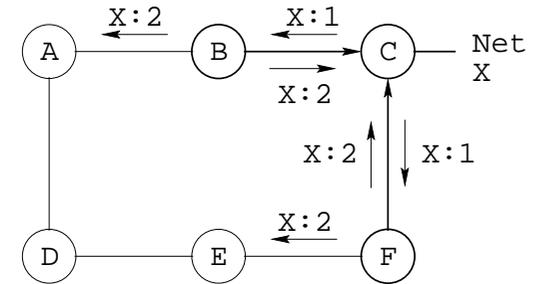
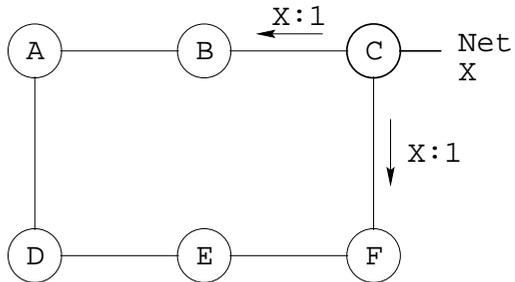
- Metric
 - 経路の質の表現
 - 1 ~ 16 の整数
 - 小さな metric が良い経路
 - 一般には hop 数を表現
 - metric 16
 - 「使用禁止」ことの表明
 - 他の経路を使用 or 届かない
 - 受信時には +n してから
 - 実際には +1 することが多い

- 同じ metric の経路
 - 両方が候補になり得る
 - RIP の場合は、どちらかを選択
 - 既に存在している経路
 - もし holddown しかけていたら
 - 新たな経路を使ってもよい

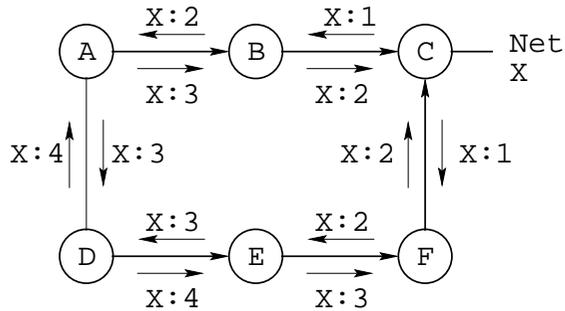


- 再現性のないアルゴリズムは良くない
 - デバッグが面倒
- インターネットでは
 - 「先にある経路を使用」
 - どっちが先かは分からない
 - OSPF や BGP4 では
 - tie breaker
 - 到着順以外のパラメータで選択
 - ある状態に収束する

- 30 秒毎に経路表をアナウンス
 - 30 秒経つと次のルータ
 - 全体に行き渡るのには時間がかかる



経路の伝搬 (3)



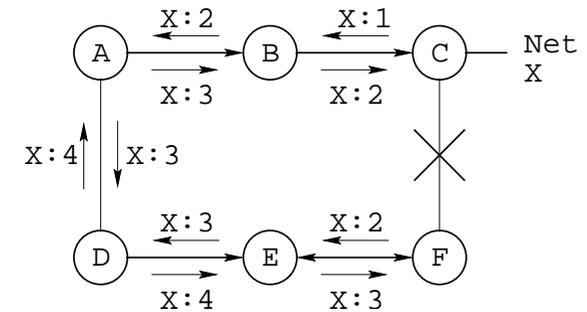
- 定常状態

- 経路が行き渡った状態
- 経路の変動がない状態
- 結構時間が掛かる
 - ネットワークが大きい場合

リンクが切れたら (1)

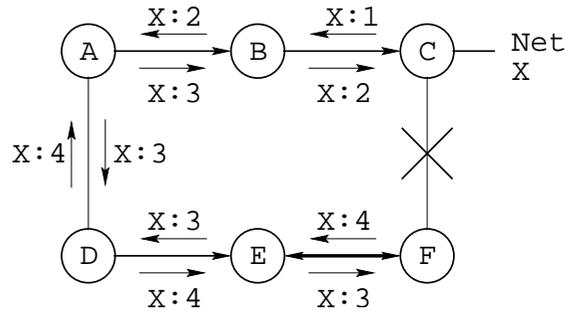
- C, F 間のリンクがダウンすると :

- F では C からの RIP が途絶える
 - 180 秒で holddown
 - metric を 16



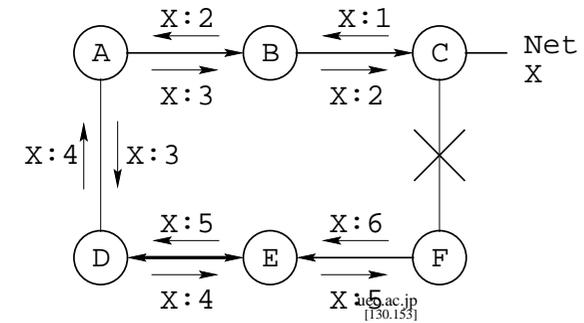
リンクが切れたら (2)

- F での X の経路が holddown になると
 - E からの経路を採用
 - 3 は 16 より良い
 - F の X 行き経路が E を向く
 - E, F 間で経路ループが発生



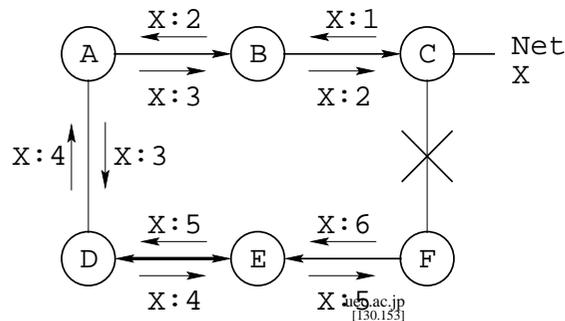
リンクが切れたら (3)

- F から E に X:6 が送られる
 - E は X 行き経路として D を選択
 - D, E 間で経路ループが発生



リンクが切れたら (4)

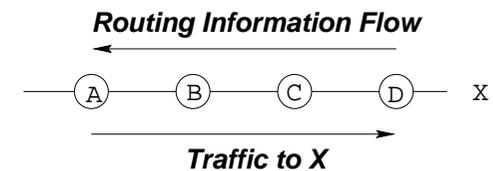
- E から D に X:5 が送られると：
 - D は X 行き経路として A を選択
 - 経路ループの無い状態に辿り着く



- ループ解消に時間が掛かる
 - 特にネットワークが大きい場合

RIP とトラフィック

- RIP 情報の伝搬
 - source X から離れる方向
- X へのトラフィック
 - X に向かう方向
- 一般には RIP とトラフィックは逆方向



RIP パケット

Dec 15, 1998
WIDE Project
Foil 40

WIDE

CMD	Vers(1)	MBZ
AFI (IP:2)		MBZ
Destination		
MBZ		
MBZ		
Metric (1-16)		

*Repeated
for
each
Destination*

MBZ: Must Be Zero

RIP コマンド

Dec 15, 1998
WIDE Project
Foil 41

WIDE

- 1: request
 - 隣接ルータに対する経路要求
 - 特定の宛先の問い合わせ
 - 経路表全体の要求
 - AFI: 0, Metric: 16
- 2: response
 - 問い合わせに対する返事
 - 30 秒毎の更新メッセージ
 - triggered update

RIP の改良 (1)

Dec 15, 1998
WIDE Project
Foil 42

WIDE

- 次段ルータからの metric 16
 - metric を 16 に設定
 - 120 秒の GC タイマを起動
 - 180 秒の holdtime を省略
- Triggered Update
 - 経路変動があった場合
 - 30 秒を待たずにアナウンス
 - 変化した経路を素早く伝搬

RIP の改良 (2)

Dec 15, 1998
WIDE Project
Foil 43

WIDE

- Split Horizon
 - 経路が来た方向にはアナウンスしない
 - 誤った経路選択を防止
 - オーバヘッドの低減
 - 無効な経路広告を抑制
- Poisoned Reverse
 - Metric 16 でアナウンスする
 - ループ解消が速い場合がある
 - オーバヘッドは増加
 - 全経路分のアナウンス

RIP 実装上の改良

Dec 15, 1998
WIDE Project
Foil 44

WIDE

- 複数の経路候補を保持
 - 最小 metric のものを使用
 - holddown 等が発生したら
 - 直ちに次善の経路を設定
 - 経路広告の待ち時間を節約
 - ほとんどの router や Gated

RIP の問題点

Dec 15, 1998
WIDE Project
Foil 45

WIDE

- 30 秒に一回の Flooding
 - 一パケットで 25 経路
 - 経路数の増加でルータの負荷が増大
 - ルータの出力・入力キュー
 - オーバフローの可能性
- Metric が 1 ~ 16
- 経路収束が遅い
 - 大きなネットワークには適応できない
- CIDR に適合しない
 - Netmask の伝達ができない

RIP と subnet

Dec 15, 1998
WIDE Project
Foil 46

WIDE

- RIP は netmask を伝搬しない
 - ルータローカルな知識で判断
- 宛先とルータの net 部が同じだったら
 - subnet 部と host 部に分割
 - host 部が non-zero : host 経路
 - host 部が all-zero : subnet 経路
- 宛先とルータの net 部が異なる場合
 - host 部が non-zero : host 経路
 - host 部が all-zero : net 経路

RIP の適用範囲

Dec 15, 1998
WIDE Project
Foil 47

WIDE

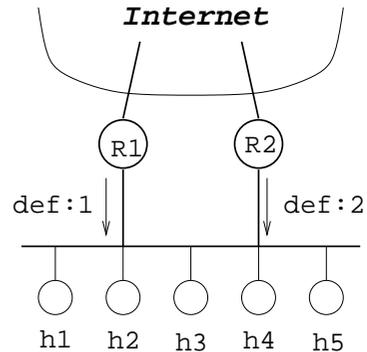
- 小さなネットワーク
 - 数十サブネット、直径で数ホップ
- ルータの存在の通知
 - default 経路のみ通知
 - ホストは受信だけ
 - routed -q
 - IPv6 では RA として実現

RIP の適用範囲 (2)

Dec 15, 1998
WIDE Project
Foil 48

WIDE

- バックアップルータへの切替え
 - default 経路のみ通知
 - 異なった metric
 - ホストでは RIP 受信
 - HSRP の方がいいか？



RIP version 2

Dec 15, 1998
WIDE Project
Foil 49

WIDE

- RIP の改良版
 - RFC1723 : Draft Standard
 - CIDR に対応
 - RIP に上位互換
 - マルチキャストの使用
 - 関係の無いホストの負荷低減
 - Metric は 1 ~ 16
 - 変更できず ...
 - アルゴリズムは変更せず

RIP-2 の特徴

Dec 15, 1998
WIDE Project
Foil 50

WIDE

- RIP に負荷情報の伝搬機能を追加
 - ”carrying additional information”
 - RIP の未使用フィールドの活用
- Netmask の搬送 : 4byte
 - CIDR に対応
- Nexthop アドレスの通知 : 4byte
 - Redirect を避ける
- Tag の添付 : 2 byte
 - BGP との連携

RIP-2 の特徴 (2)

Dec 15, 1998
WIDE Project
Foil 51

WIDE

- 認証機能
 - メッセージ単位
 - MBZ だけでは収まらない
 - 1 経路分の場所を専有
 - AFI: 0xffff
 - パスワード認証のみ定義
 - 非暗号化パスワード
 - 最大 16byte

RIP-2 のパケット

Dec 15, 1998
WIDE Project
Foil 52

WIDE

CMD	Vers(2)	unused
AFI (IP:2)		Route TAG
Destination		
Subnet Mask		
Next Hop		
Metric (1-16)		

Repeated for each Destination

RIP-2 の認証パケット

Dec 15, 1998
WIDE Project
Foil 53

WIDE

CMD	Vers(2)	unused
0xffff		AuthType
16byte auth data		
AFI (IP:2)		Route TAG
Destination		
Subnet Mask		
Next Hop		
Metric (1-16)		

Authentication Header

Repeated for each Destination

RIP-2 の適用

Dec 15, 1998
WIDE Project
Foil 54

WIDE

- 実装 :
 - gated, cisco, bay 等
- 実際にはあまり嬉しくない
 - OSPF が普及、安定に稼働
 - ホストでの実装が少ない
 - Default だけなら RIP-1 で充分
 - DHCP による default 経路の設定
 - 30 秒に 1 回の割り込み
 - あまり苦にならない

OSPF

OSPF

Dec 15, 1998
WIDE Project
Foil 56

WIDE

- Open Shortest Path First
 - RFC2328 で定義
 - 210 page を超える厚さ !
- Standard Protocol : STD0054
 - RFC1131 : 1989 Oct, PS
 - RFC1247 : 1991 Jul, DS
 - RFC1583 : 1994 Mar, DS
 - RFC2178 : 1997 Jul, DS
 - RFC2328 : 1998 Apr, S

OSPF の特徴

Dec 15, 1998
WIDE Project
Foil 57

WIDE

- リンク状態型経路制御
 - LSA の生成
 - 各ルータ、各ネットワーク
 - LSA データベースの同期機構
 - Adjacency を利用
 - エリアによる階層構造
 - エリア境界での経路の集約
 - オーバヘッドの低減
 - 外部経路の取扱い
 - AS External 経路

OSPF の特徴 (2)

Dec 15, 1998
WIDE Project
Foil 58

WIDE

- プロトコル
 - IP プロトコル番号 89
 - 独自の再送プロトコル
 - ケーブルマルチキャスト
 - 再送はユニキャスト
 - 認証によるセキュリティ
 - 単純パスワード
 - 共有秘密鍵と MD5

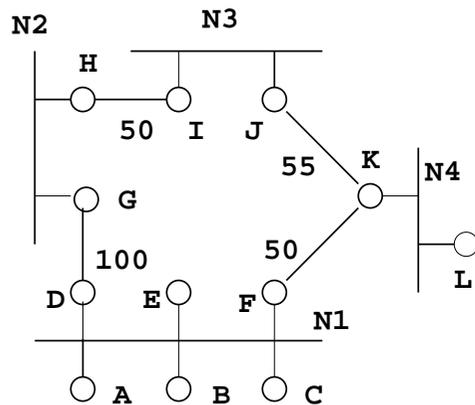
リンク状態型経路制御

Dec 15, 1998
WIDE Project
Foil 59

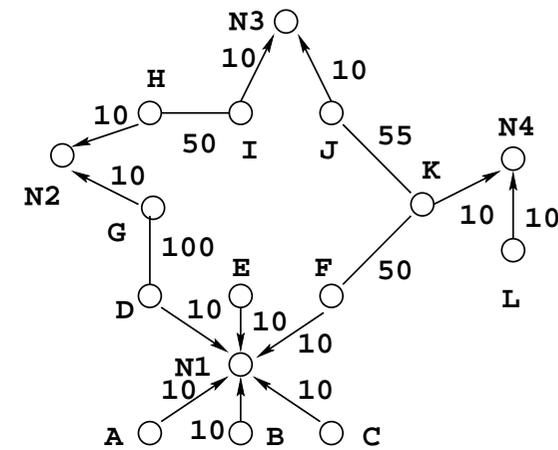
WIDE

- トポロジデータベース
 - 沢山の LSA から構成
- 各ルータで同一コピーを保持
 - 信頼性のある通信が必要
 - 変動を実時間に更新
- 共通なデータベースから経路計算
 - 共通なアルゴリズムを使用
 - そうしないとループが発生
 - 経路ループは短時間で解消
 - ポリシの実現は困難

- Dijkstra のアルゴリズム

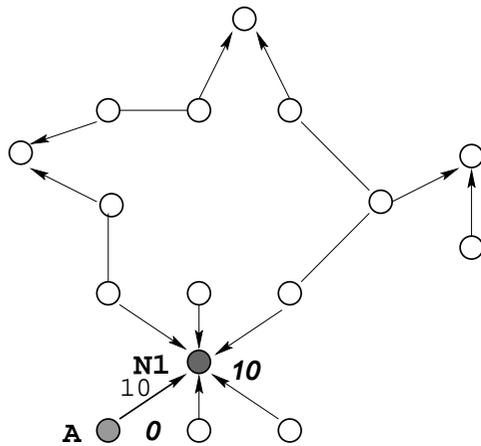


- マルチアクセスネットワークもノード
- 出力側に metric (入力側は metric 0)



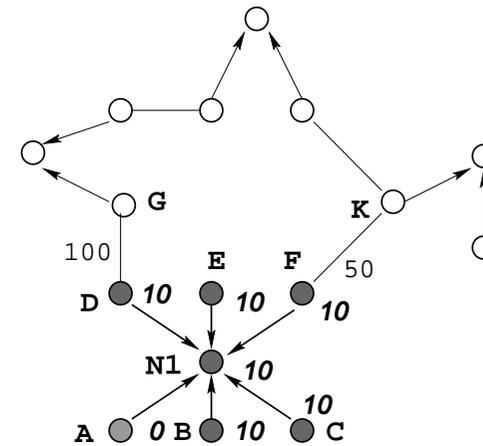
経路の計算 (3)

- ノード A に対する SPF 木の計算
 - A から N1 へ



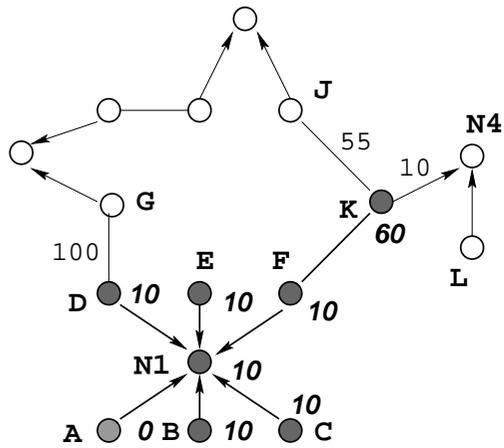
経路の計算 (4)

- ノード A に対する SPF 木
 - N1 の隣接ノード



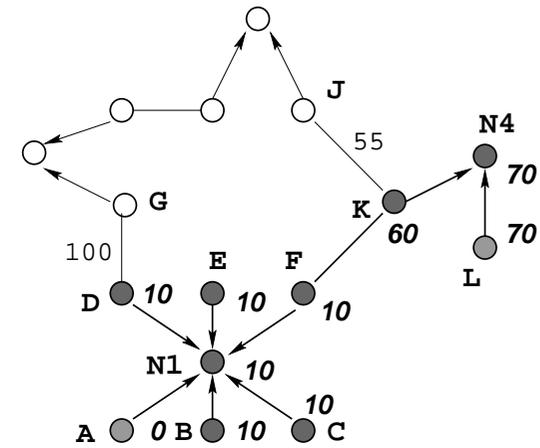
経路の計算 (5)

- ノード A に対する SPF 木
 - K を加える



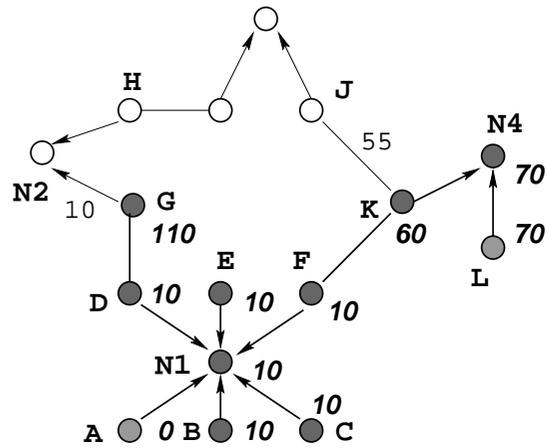
経路の計算 (6)

- ノード A に対する SPF 木
 - N4, L を加える



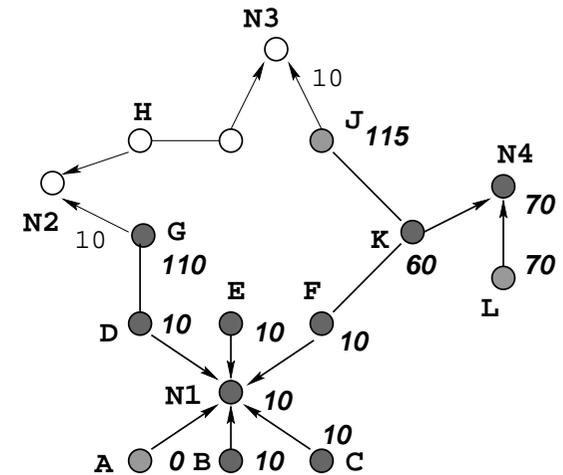
経路の計算 (7)

- ノード A に対する SPF 木
 - G を加える



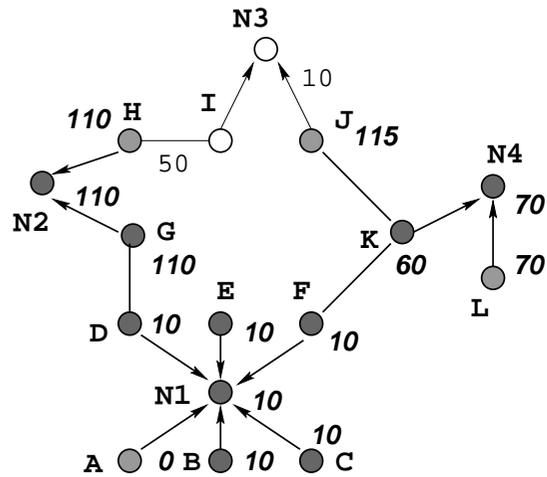
経路の計算 (8)

- ノード A に対する SPF 木
 - J を加える



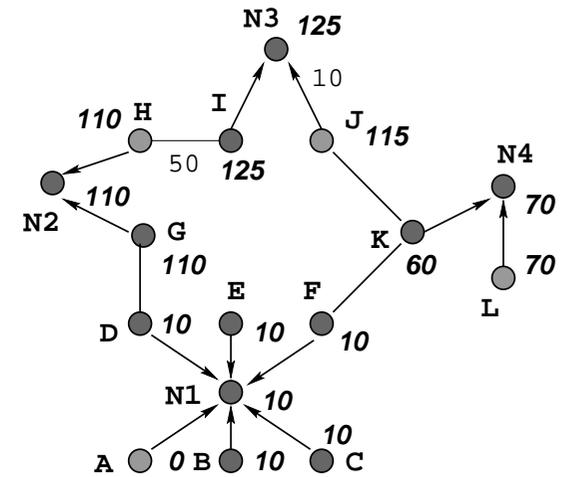
経路の計算 (9)

- ノード A に対する SPF 木
 - N2, H を加える



経路の計算 (10)

- ノード A に対する SPF 木
 - N3, I を加える

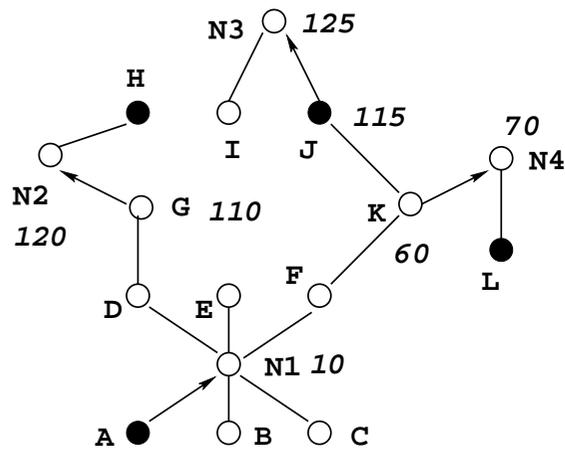


経路の計算 (13)

Dec 15, 1998
WIDE Project
Foil 70

WIDE

- 完成した SPF 木



経路の計算 (12)

Dec 15, 1998
WIDE Project
Foil 71

WIDE

- ノード A に対する経路表
 - G : 110 (NH: D)
 - K : 60 (NH: F)
 - J : 115 (NH: F)
 - N1 : 10 (NH: A)
 - N2 : 120 (NH: D)
 - N3 : 125 (NH: F)
 - N4 : 70 (NH: F)

経路計算のまとめ

Dec 15, 1998
WIDE Project
Foil 72

WIDE

- ネットワークの地図を作る
 - 作成方法は後述
 - 全ノードにコピー
- 各ノードで
 - 自分を根とする SPF 木を計算
 - 経路表を設定
- 全ノードが同じ地図から経路計算
 - 地図の同期が重要
 - 一時的な経路ループは早期に解消

データベースの構成

Dec 15, 1998
WIDE Project
Foil 73

WIDE

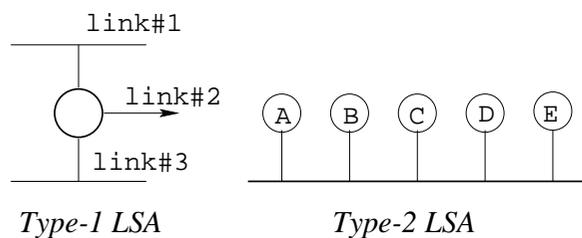
- トポロジの記述
 - Link State Advertisement の集合
 - 5 種類の LSA
 - Type-1 ~ Type-5
- ルータの名前
 - Router ID (32bit)
- ネットワークの名前
 - 指定ルータ (DR) のアドレス

データベースの構成 (2)

Dec 15, 1998
WIDE Project
Foil 74

WIDE

- Type-1 LSA : ルータの記述
 - 各ルータが生成
 - 接続ネットワークのリスト
- Type-2 LSA : ネットワークの記述
 - 各ネットワークで生成
 - 接続ルータのリスト



データベースの構成 (3)

Dec 15, 1998
WIDE Project
Foil 75

WIDE

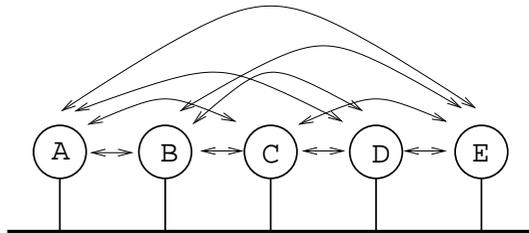
- Type-3 LSA
 - サマリ情報の記述
- Type-4 LSA
 - AS 境界ルータの記述
- Type-5 LSA
 - AS 外部経路の記述

LSA の伝搬

Dec 15, 1998
WIDE Project
Foil 76

WIDE

- 信頼性のある伝搬
- 効率的な伝搬



- ルータ数の二乗のコストは避けたい

マルチキャストの利用

Dec 15, 1998
WIDE Project
Foil 77

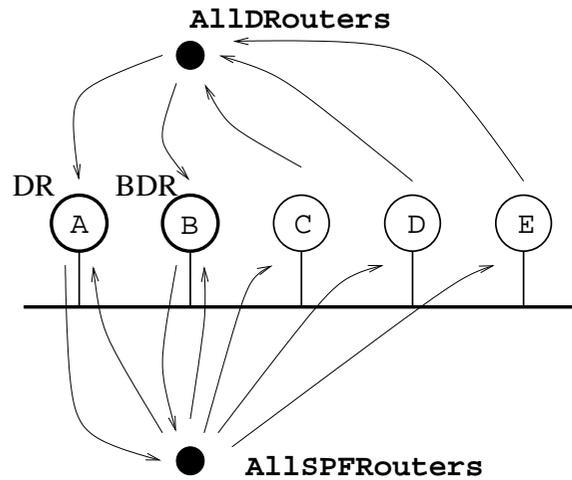
WIDE

- DR/BDR
 - Designated Router : 指定ルータ
 - Backup BDR : バックアップ指定ルータ
- ルータの関係
 - DR (BDR) とその他のルータ : Adjacent
 - それ以外 : Neighbor
- 2つのマルチキャストアドレス
 - AllSPFRouters : 全 OSPF ルータ
 - AllDRouters : DR/BDR

DR/BDR との関係

Dec 15, 1998
WIDE Project
Foil 78

WIDE



DR の役割

Dec 15, 1998
WIDE Project
Foil 79

WIDE

- Adjacency の一端
- 信頼性のあるデータ転送
- 各ルータから DR/BDR への通信
 - AllDRouters : 224.0.0.6
- DR から各ルータへの通信
 - AllSPFRouters: 224.0.0.5
- 再送
 - 常に Unicast を使用

Router ID

Dec 15, 1998
WIDE Project
Foil 80

WIDE

- 32bit の値
- 各ルータの対する ID
 - ドメイン内で一意でなければならない
 - 複数のルータが同一 ID を用いると大混乱
 - IP アドレスのいずれかを利用
- なるべく変動しないもの
 - 安定なインターフェース
 - Software インターフェース

データベース同期の概要

Dec 15, 1998
WIDE Project
Foil 81

WIDE

- Down : 未活性化状態
 - Hello の交換
- Init : Hello を受信
 - 2-Way あるいは ExStart へ
- 2-Way : 双方向の通信の確認
 - 必要なら DR / BDR の選出
 - DR / BDR 以外は最終状態
 - DR / BDR は ExStart へ
- ExStart : DD の初期交換
 - Master / Slave の決定
 - Seq no. の交換

- Exchange :
 - DD の交換
 - Master が主導
- Loading : LSA の交換
 - 必要な LSA を要求
 - LS Request パケット
 - 要求された LSA を供給
 - LS Update パケット
- Full:
 - 同期が取れた状態

- 交換される情報
 - 自分の Router ID
 - Area ID
 - ネットマスク
 - Hello Interval
 - Router Dead Interval
 - Router Priority
 - DR / BDR
 - 既知のルータの Router ID のリスト

Hello プロトコル (2)

Dec 15, 1998
WIDE Project
Foil 84

WIDE

- Hello Interval 毎に送出
 - default では 10 秒に 1 回
- 各種パラメータの確認
 - Area ID
 - ネットマスク
 - Hello Interval
 - Router Dead Interval
 - Stub Area かどうか

Hello プロトコル (3)

Dec 15, 1998
WIDE Project
Foil 85

WIDE

- 受信した Router ID のリストを添付
 - 通信の双方向性を確認
- DR / BDR を決定
 - 既存の DR / BDR を優先
 - 選択の安定性
 - Priority の大きなルータ
 - Router ID の大きなルータ
 - Priority 0 : DR にならない

Hello プロトコル (4)

Dec 15, 1998
WIDE Project
Foil 86

WIDE

- Dead Interval 内に Hello が来ないと
 - ダウンしたものと認定
 - 必要なら DR/BDR の選出
 - BDR が DR になる
 - 新 BDR を選出
- RouterDeadInterval
 - HelloInterval の数倍
- 推奨値 : Ethernet では
 - HelloInterval : 10 秒
 - RouterDeadInterval : 40 秒

データベースの同期

Dec 15, 1998
WIDE Project
Foil 87

WIDE

- Adjacent ルータ間で同期
 - ルータと指定ルータとの間
 - p2p リンクの両端
- Master / Slave の決定
 - ExStart 状態
 - Seq No. を適当に決める
 - Init / More / Master bit をセット
 - Router ID の大きい方が master
 - Master bit を off

データベースの同期 (2)

Dec 15, 1998
WIDE Project
Foil 88

WIDE

- Database Description を交換
 - Exchange 状態
 - LSA ヘッダを送信
 - 保持していない LSA を確認
 - 古い LSA を確認
- Database Description パケット
 - Seq no.
 - 最後に受信した Seq no.
 - 複数の LSA ヘッダ
 - Window Size : 1
 - More bit で継続を指示

データベースの同期 (3)

Dec 15, 1998
WIDE Project
Foil 89

WIDE

- LSA の交換
 - DD によって更新が必要な LSA を確認
 - LS Request パケットで LSA を要求
 - LS Update パケットで応答
 - 複数の LSA を一度に交換
- LSA の指定
 - LS type, LS ID, 広告ルータ ID
 - 最新のものを送信

Serial リンクや NBMA

Dec 15, 1998
WIDE Project
Foil 90

WIDE

- Serial リンク
 - 両端同士でデータベースの同期
 - Priority は無意味
- NBMA
 - 各ルータに DR/BDR 候補を設定
 - HelloInterval : 30 sec
 - RouterDeadInterval : 120sec
 - Down している neighbor にも Hello
 - PollInterval : 120 sec

共通パラメータ

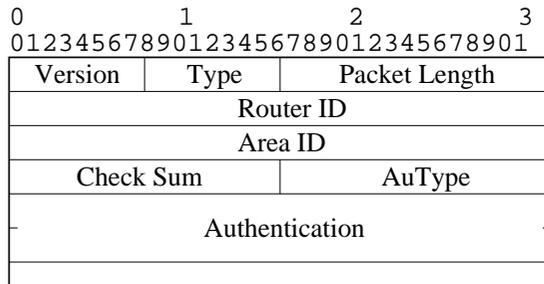
Dec 15, 1998
WIDE Project
Foil 91

WIDE

- ネットワークで共通
 - HelloInterval
 - RouterDeadInterval
 - Network Mask
 - 認証データ
- エリアで共通
 - 認証方式
 - エリア ID
 - E bit
 - stub エリアかどうか

OSPF パケットヘッダ

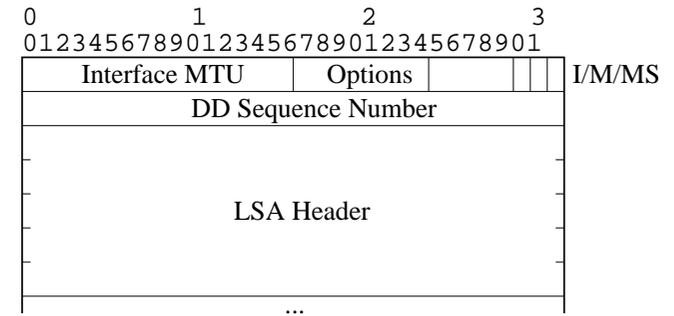
Dec 15, 1998
WIDE Project
Foil 92 **WIDE**



- Type:
 - 1 - Hello
 - 2 - Database Description
 - 3 - LS Request
 - 4 - LS Update
 - 5 - LS Acknowledgment

DD パケット

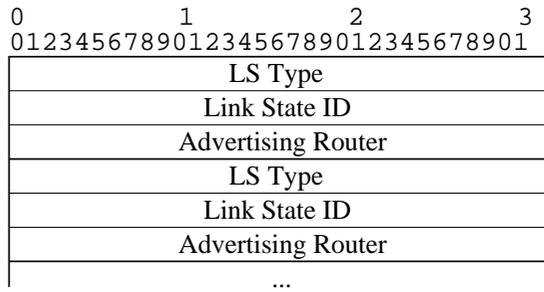
Dec 15, 1998
WIDE Project
Foil 93 **WIDE**



- 複数の LSA ヘッダを収容

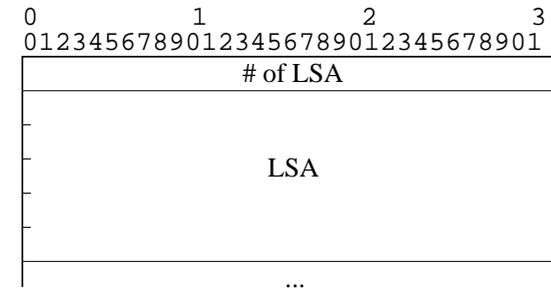
LS Req パケット

- LSA を要求
 - LS Type, LS ID, Adv Rtr



LS Update パケット

- LSA の更新を通知
 - 複数の LSA

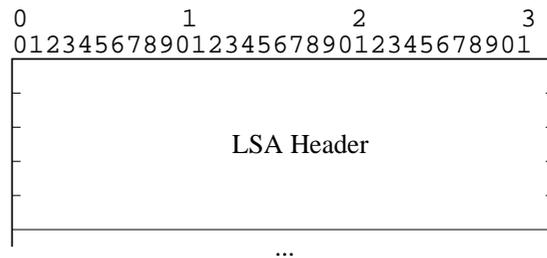


LS Ack パケット

Dec 15, 1998
WIDE Project
Foil 96

WIDE

- LS Update を ack



LSA

Dec 15, 1998
WIDE Project
Foil 97

WIDE

- Link State Advertisement
 - LSA Type, LS Age
 - Link State ID
 - Sequence Number
 - 0x80000000 ~ 0x7fffffff
 - 一周したら特別な処理
 - LS checksum, LS length
 - Expire Time
 - 1800 sec で再生成
 - 3600 sec でタイムアウト

LSA の生成

Dec 15, 1998
WIDE Project
Foil 98

WIDE

- Router LSA

- Type 1 LSA
- 各 OSPF ルータが生成
- LSID : 生成したルータの Router ID
- 各接続ネットワークについて
 - Link ID : 隣接リンクの ID 等
 - Link Data : Link の IP アドレス等
 - 外向きのコスト (TOS 毎に)

LSA の生成 (2)

Dec 15, 1998
WIDE Project
Foil 99

WIDE

- ネットワーク LSA

- Type 2 LSA
- マルチアクセスネットワークに対応
 - 2 台以上のルータが接続したもの
 - 一台だけの場合には、stub network
- DR によって生成
- LSID : DR の IP アドレス
- ネットワークマスク
- 接続ルータの Router ID のリスト

Metric

Dec 15, 1998
WIDE Project
Foil 100

WIDE

- リンクの出力側に metric
 - Ethernet : ルータ毎に設定
 - Serial : 方向別に設定
 - 一方向リンクは対称外
 - 近接関係を確立できない
 - UDLR working group で作業中
 - 入力側のコストは 0 と考える

Metric (2)

Dec 15, 1998
WIDE Project
Foil 101

WIDE

- 1 ~ 65535 の範囲
- 推奨値 : 10E8 / speed
 - FDDI : 1
 - Ethernet : 10
 - 1.5Mbps : 66
 - 64kbps : 1562
- 気にしない方がいいような ?!

Metric (3)

Dec 15, 1998
WIDE Project
Foil 102

WIDE

- Equal Cost Multipath
 - 同じコストの異なる経路
 - 負荷分散が可能
 - ルータの転送方式に依存
 - per packet
 - per destination

エリア

Dec 15, 1998
WIDE Project
Foil 103

WIDE

- AS (経路ドメイン) をエリアに分割
 - 32bit エリア ID
 - 0.0.0.0 はバックボーン
 - 各ネットワークは単一エリアに所属
 - エリア境界はルータ
 - エリア外部はサマリ情報
 - エリア内だけ Dijkstra で経路計算
 - オーバヘッドの低減
 - サマリによる経路の集約

エリア (2)

Dec 15, 1998
WIDE Project
Foil 104

WIDE

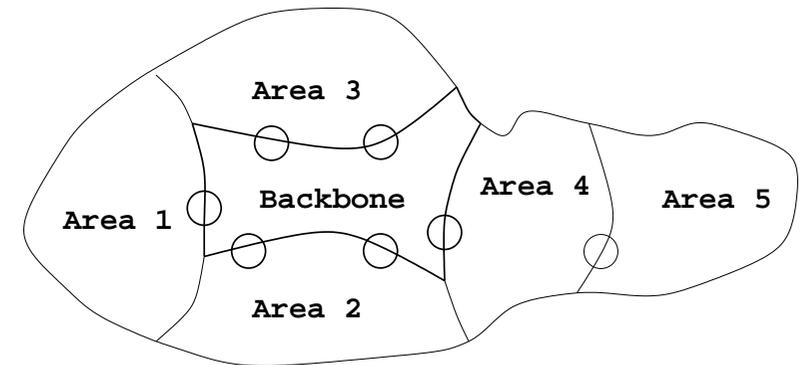
- エリア間トラフィック
 - 発信元エリアから
 - バックボーンエリアを経由
 - 宛先エリアに
- バックボーンエリア
 - 全てのエリアに隣接
 - 必要な場合には Virtual Link で接続

エリア (3)

Dec 15, 1998
WIDE Project
Foil 105

WIDE

- 各エリアは連結・バックボーンに隣接
- 正しくないエリア分割の例

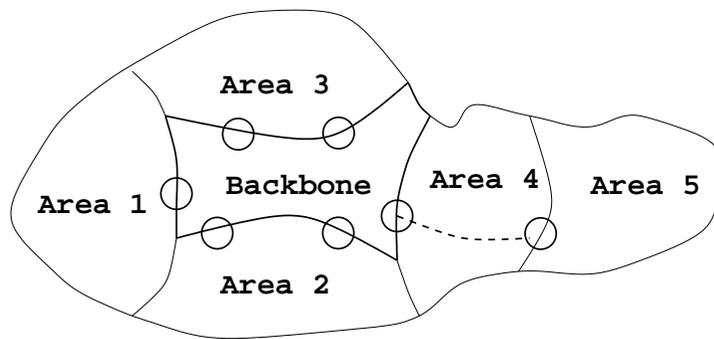


エリア (4)

Dec 15, 1998
WIDE Project
Foil 106

WIDE

- Virtual Link で接続
 - バックボーンを拡張
 - 単一エリアを横断



ルータの種類

Dec 15, 1998
WIDE Project
Foil 107

WIDE

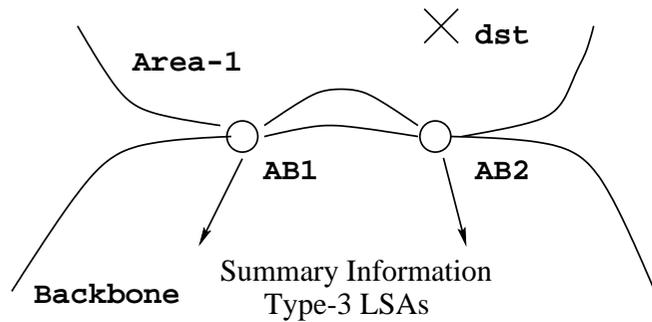
- 内部ルータ
 - 単一エリアのみに接続
- バックボーンルータ
 - バックボーンに属しているもの
 - 内部ルータまたはエリア境界ルータ
- エリア境界ルータ
 - バックボーンと他のエリアを接続
- AS 境界ルータ
 - AS と他の AS を接続
 - Type-5 LSA を生成

エリア境界ルータ

Dec 15, 1998
WIDE Project
Foil 108

WIDE

- Type-3 LSA : サマリ情報の伝搬
 - エリア外部かつ AS 内の各宛先
 - LSID: 宛先アドレス
 - 各エリア境界ルータで生成



外部の経路

Dec 15, 1998
WIDE Project
Foil 109

WIDE

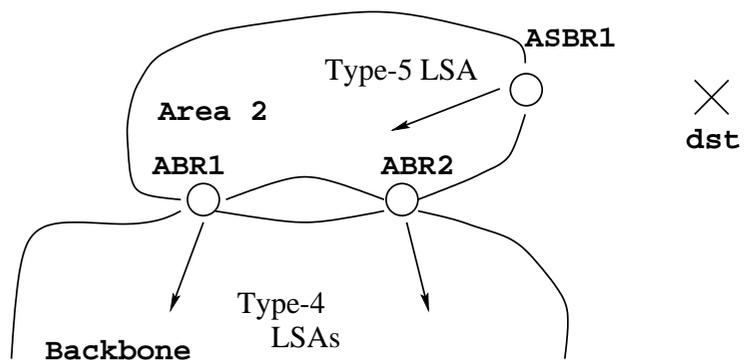
- エリア境界ルータ
 - Type-4 LSA も生成
 - AS 境界ルータのサマリ情報
 - AS 境界ルータへのコストを表現
 - LSID: AS 境界ルータの Router ID
- AS 境界ルータ
 - Type-5 LSA の生成
 - AS 外部の各宛先について一つ
 - LSID: 宛先アドレス

外部の経路 (2)

Dec 15, 1998
WIDE Project
Foil 110

WIDE

- Type 4 LSA
 - 各 ASBR について ABR が生成
- Type 5 LSA
 - 各外部宛先について ASBR が生成



Type 5 LSA

Dec 15, 1998
WIDE Project
Foil 111

WIDE

- Metric Type 1
 - 外部 metric と内部 metric は比較可能
 - metric の合計が小さい経路を選択
- Metric Type 2
 - 外部 metric が支配的と考える
 - 外部 metric の小さな経路を選択
 - 内部 metric は考えず

Type1 LSA



	0					1						2						3							
	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	
V/E/B	0										0														# of links
	Link ID																								
	Link Data																								
	Type	# of TOS					metric																		
	TOS	0					TOS metric																		

Type1 LSA (2)



- type 1 : p2p
 - LID: 隣の Router ID
- type 2 : transit network
 - LID: DR のアドレス
- type 3 : stub network
 - LID: (サブ) ネットワーク部
- tyoe 4 : virtual link
 - LID: 隣の Router ID
- Link Data : 自分のアドレス

Type1 LSA (3)

Dec 15, 1998
WIDE Project
Foil 116

WIDE

- bit V
 - Virtual link の片端
- vit E
 - AS External ルータかどうか
- vit B
 - エリア境界ルータかどうか

Type2 LSA

Dec 15, 1998
WIDE Project
Foil 117

WIDE

- ネットワーク
 - 接続しているルータ一覧
 - DR が生成

0	1	2	3
0	1	2	3
1	2	3	0
2	3	0	1
3	0	1	2
Network Mask			
Attached Router			
...			

Type 3/4 LSA

Dec 15, 1998
WIDE Project
Foil 118

WIDE

- サマリ情報

- Type-3 IP ネットワーク
- Type-4 ASBR

0																1																2																3															
012345678901234567890123456789012345678901																																																															
Network Mask																																																															
0																metric																																															
TOS																TOS metric																																															

Type 5 LSA

Dec 15, 1998
WIDE Project
Foil 119

WIDE

- AS 外部の経路

- E=0 : type-1
- E=1 : type-2

0																1																2																3															
012345678901234567890123456789012345678901																																																															
Network Mask																																																															
E 0																metric																																															
Forwarding Address																																																															
External Route Tag																																																															
E TOS																TOS metric																																															
Forwarding Address																																																															
External Route Tag																																																															

Stub エリア

Dec 15, 1998
WIDE Project
Foil 120

WIDE

- Type-5 LSA を転送しない
- 全てのエリア内 OSPF ルータの合意が必要
- Default 経路を利用
 - エリア境界ルータで生成
 - Type-3 LSA で伝搬
- 外部経路の取り扱いをしない
 - Type-5 LSA を送らない
 - 例外 :NSSA オプション

認証

Dec 15, 1998
WIDE Project
Foil 121

WIDE

- Routing Protocol
 - 経路制御のキー
 - Router は host と同程度に secure
 - 認証機構は必要 (暗号化は不要)
- 無認証
- パスワード認証
 - 平文パスワード
 - Replay し放題

- 暗号化認証

- Seq No. が一周すると
 - Replay 可能
- 暗号化 Seq No. を導入
 - sec from boot など
 - 当分一周しない
- Key ID
 - 複数の Key を利用可
 - Key の取り換え時に便利
- MD5 署名をパケットの末尾に
 - 16byte

- LSID 規則の緩和

- RFC2328 Appendix E
- Type-3, Type-5 LSA
- マスクの異なる同一宛先を表現
 - ホスト部ビットを使用
 - 異なった LSID を命名

- Router LSA に追加
 - 自分の i/f を追加
 - stub network
 - Link ID: 自分のアドレス
 - Link Data: 0xffffffff
 - 各隣接サイトにつき、下記を追加
 - Point-to-point link
 - Link ID: 相手の Router ID
 - Link Data: 自分のアドレス
 - Frame-Relay などに便利

OSPF の運用

OSPF の設定例

Dec 15, 1998
WIDE Project
Foil 126

WIDE

- インターフェースの設定

```
interface ethernet 0/0
  ip address 192.168.1.1 255.255.255.0
  ip ospf authentication-key himitsu
  ip ospf cost 10
  ip ospf priority 1
  ip ospf hello-interval 10
  ip ospf dead-interval 40
```

OSPF の設定例 (2)

Dec 15, 1998
WIDE Project
Foil 127

WIDE

- OSPF プロトコルの設定

```
router ospf 100
  network 192.168.1.1 0.0.0.0 area 0.0.0.0
  area 0.0.0.0 authentication
```

Interface の確認

Dec 15, 1998
WIDE Project
Foil 128

WIDE

```
% show ospf interface
Ethernet0/0/0 is up, line protocol is up
  OSPF not enabled on this interface
Fddi5/0/0 is up, line protocol is up
  Internet Address 203.178.137.34/29, Area 0.0.0.0
  Process ID 2500, Router ID 203.178.136.4, Networ
  Transmit Delay is 1 sec, State BDR, Priority 1
  Designated Router (ID) 203.178.136.6, Interface
  Backup Designated router (ID) 203.178.136.4, Int
  Timer intervals configured, Hello 10, Dead 40, W
  Hello due in 00:00:07
Neighbor Count is 1, Adjacent neighbor count is
  Adjacent with neighbor 203.178.136.6 (Designa
```

Neighbor の確認

Dec 15, 1998
WIDE Project
Foil 129

WIDE

```
% show ospf neighbor
```

Neighbor ID	Pri	State	Dead Time	Address
203.178.136.6	1	FULL/DR	00:00:35	203.178.137.33
150.65.0.1	1	FULL/ -	00:00:35	203.178.136.126
203.178.141.74	1	FULL/ -	00:00:38	203.178.141.74
203.178.136.7	1	FULL/ -	00:00:32	203.178.141.66
203.178.141.94	1	FULL/ -	00:00:36	203.178.141.94
203.178.141.85	1	FULL/ -	00:00:30	203.178.141.82

Neighbor の確認 (2)

Dec 15, 1998
WIDE Project
Foil 130

WIDE

```
% show ospf neighbor
```

Neighbor ID	Pri	State	Dead Time	Address
203.178.136.195	1	2WAY/DROTHER	00:00:55	203.178.136.167
203.178.137.50	1	FULL/BDR	00:00:56	203.178.136.164
203.178.136.196	1	2WAY/DROTHER	00:00:59	203.178.136.165
203.178.136.169	1	2WAY/DROTHER	00:00:55	203.178.136.169
203.178.136.241	1	FULL/DR	00:00:59	203.178.136.170
203.178.136.168	1	2WAY/DROTHER	00:00:55	203.178.136.168

Neighbor の確認 (3)

Dec 15, 1998
WIDE Project
Foil 131

WIDE

```
% show ip ospf neighbor detail
```

```
Neighbor 203.178.136.196, interface address 203.178.136.101
  In the area 0.0.0.0 via interface Serial1/0
  Neighbor priority is 1, State is FULL
  Options 2
  Dead timer due in 00:00:59
Neighbor 203.178.142.254, interface address 203.178.142.254
  In the area 203.178.136.100 via interface Serial1/3
  Neighbor priority is 1, State is FULL
  Options 2
  Dead timer due in 00:00:57
  Link State retransmission due in 00:00:04
    LSA in retransmission queue 2
```

Neighbor にならない場合

Dec 15, 1998
WIDE Project
Foil 132

WIDE

- パラメータの不一致
 - エリア ID
 - Netmask
 - タイミングパラメータ
 - 認証方式
 - 認証
- 状況の確認

```
# debug ip ospf adj
```

Database の確認

Dec 15, 1998
WIDE Project
Foil 133

WIDE

```
% show ip ospf database database-summary
```

Area ID	Rtr	Net	SumNet	SumASBR	Subtotal
0.0.0.0	61	17	46	48	172
203.178.136.100	2	0	97	70	169
AS External					4452
Total	63	17	143	118	4793

Database の確認 (2)

Dec 15, 1998
WIDE Project
Foil 134

WIDE

```
% show ip ospf database
```

```
Router Link States (Area 0.0.0.0)
```

Link ID	ADV Router	Age	Seq#	Checksum	Link
133.27.2.3	133.27.2.3	340	0x80000A44	0x11F9	1
150.65.0.1	150.65.0.1	999	0x8000613E	0xEEF4	5
202.249.3.84	202.249.3.84	622	0x800015B4	0x53C0	1
202.249.3.86	202.249.3.86	666	0x80001638	0x2760	1
203.178.136.1	203.178.136.1	961	0x80004965	0x65D6	3
203.178.136.2	203.178.136.2	1201	0x8000144A	0x35EB	2
203.178.136.3	203.178.136.3	690	0x800054DA	0x7682	3
203.178.136.4	203.178.136.4	949	0x800018D2	0x1CCB	11

Database の確認 (3)

Dec 15, 1998
WIDE Project
Foil 135

WIDE

```
% show ip ospf database network 203.178.137.1
```

```
Routing Bit Set on this LSA
```

```
LS age: 854
```

```
Options: (No TOS-capability)
```

```
LS Type: Network Links
```

```
Link State ID: 203.178.137.1 (address of DR)
```

```
Advertising Router: 203.178.136.150
```

```
LS Seq Number: 8000001E
```

```
Checksum: 0x36B2
```

```
Length: 32
```

```
Network Mask: /29
```

```
Attached Router: 203.178.136.150
```

```
Attached Router: 203.178.136.16
```

Database の確認 (4)

Dec 15, 1998
WIDE Project
Foil 136

WIDE

```
% show ip ospf database database-summary
```

```
% show ip ospf database
```

```
% show ip ospf database router <LSID>
```

```
% show ip ospf database network <LSID>
```

```
% show ip ospf database summary <LSID>
```

```
% show ip ospf database asbr-summary <LSID>
```

```
% show ip ospf database external <LSID>
```

OSPF の設定ミスの例

Dec 15, 1998
WIDE Project
Foil 137

WIDE

- エリア分割
 - バックボーンに隣接しないエリアの作成
 - エリア分割の見直し
 - Virtual Link の使用
 - 単一エリアのルータ数
 - 数十程度以下が推奨
- Full Route
 - 多くのルータは 55,000 経路の OSPF は無理
 - メモリや CPU の問題
 - トラフィック

OSPF の設定ミスの例 (2)

Dec 15, 1998
WIDE Project
Foil 138

WIDE

- 同じ Router ID を複数のルータ
 - 誤った経路計算
 - ループの発生
 - <<Sequence Number の急上昇!>>
 - 対策
 - Router ID を訂正
 - 一時間待つ

OSPF tips

Dec 15, 1998
WIDE Project
Foil 139

WIDE

- 単一エリアに属すルータ
 - 全部のインターフェースを一つのエリアに

```
router ospf 100
network 0.0.0.0 255.255.255.255 area <AREA>
```

- 他にルータの無いネットワーク
 - Hello を止める

```
router ospf 100
passive-interface ethernet 0/0
```

- OSPF 以外の経路の導入

```
router ospf 100
  redistribute rip route-map rip-ospf subnet
access-list 1 permit any
route-map rip-ospf permit 10
  match ip address 1
  set metric 1000
  set metric-type type-1
```

- 外部経路の再導入

```
# clear ip ospf redistribute
```

まとめ

RIP と OSPF

Dec 15, 1998
WIDE Project
Foil 142

WIDE

- RIP
 - 限られた場面では依然として有効
 - 大きいネットワークでは使用禁止
- OSPF
 - プロトコルは複雑
 - 実装は概ね安定している
 - 運用はそんなに難しくない
 - デバッグはちょっと大変？

参考文献

Dec 15, 1998
WIDE Project
Foil 143

WIDE

- Interconnections
 - Radia Perlman, Addison-Wesley,
 - ISBN 0-201-56332-1
 - 和訳 :加藤監訳 Interconnections,
 - ソフトバンク、ISBN 4-89052-712-5
- Internet Routing Architectures
 - Bassam Halabi, Cisco Press,
 - ISBN 1-56205-652-2

参考文献 (2)

Dec 15, 1998
WIDE Project
Foil 144

WIDE

- Routing in the Internet
 - Cristian Huitema, Prentice Hall
 - ISBN 0131321927
- Managing IP networks
 - with Cisco Routers
 - Ballew, et al, O'Reilly
 - ISBN 1565923200
- OSPF: Anatomy of an Internet
 - Routing Protocol
 - John Moy, Addison-Wesley,
 - ISBN 0201634724

参考文献 (3)

Dec 15, 1998
WIDE Project
Foil 145

WIDE

- Cisco Router OSPF
 - - Design and Implementation Guide -
 - William R. Parkhurst, McGraw Hill
 - ISBN 0070486263
- OSPF Network Design Solutions
 - Thomas M., II Thomas, Cisco Systems
 - ISBN 1578700469