SONY

Ethernet LANの憂鬱

平賀十志男
ソニーグローバルソリューションズ

JANOG17 / 2006.1.20

Sony Global Solutions

はじめに

SONY

- LANを構築する場合Ethernetを使用することが多い
- EthernetでLANを構築してみるとさまざまな問題が起こることがある
  - ループに関する問題は多いがそれ以外もある
  - Ethernetにまつわる仕様、登場人物の多さ、相互接続性
- 大規模LAN構築のなかで起こりがちないくつかの問題を見ながら、これからのEthernetのあるべき姿を議論したい
  - ああして欲しい、こうして欲しい
  - あれが欲しい、これが欲しい

Sony Global Solutions

VLANを使う大規模LAN

SONY

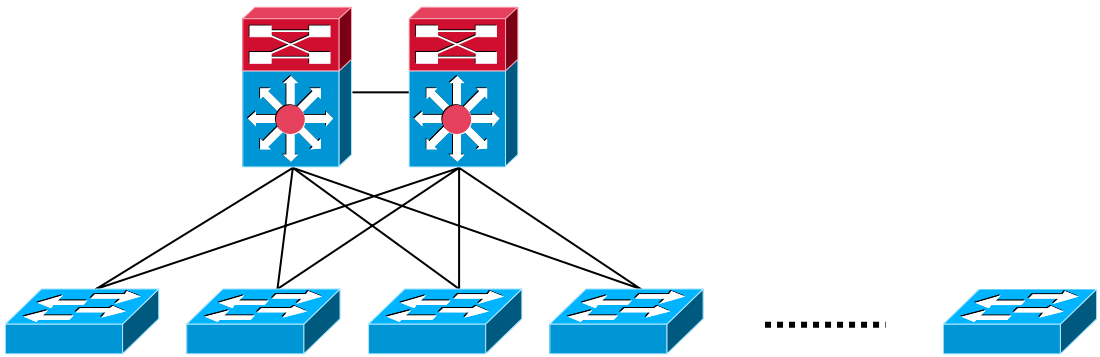
- Motivation
  - **物理トポロジーに制約されない柔軟な論理セグメンテーションを行いたい**
- Situation
  - **比較的大規模(接続ノードは数千台規模)**
  - **エッジスイッチは数十台から数百台程度接続する**
  - **接続ノードは細かくセグメンテーションされる**
  - SLBなどVLANを大量に消費する機器もある
- Requirement
  - **障害に備えて冗長化したい**
  - VLAN数は規格の上限まで使用したい
    - IEEE802.1Q → ***4094*** (VID 0 and FFF are reserved)

Sony Global Solutions

考慮点: スパニングツリー

SONY

- スパニングツリーはどうすれば…



Sony Global Solutions

考慮点: VPI (Vlan Port Instances)

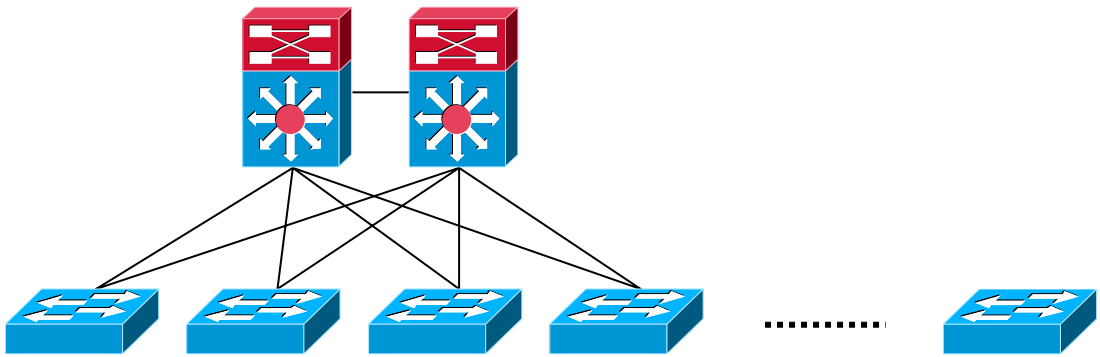
SONY

- VPIとはスイッチ全体・モジュールごとのスケーラビリティを表す指標
  - C社機器での構成設計時には必ず考慮すべき概念
- VPI数とは機器が管理する論理的なポート数の合計
  - VPI数 = Trunkポート数 × Trunk内のVLAN数 + Non-Trunkポート数
  - VPI数が増えるとスイッチ単体でSTPの処理をする論理ポート数が増えるため、BPDUの送受信およびトポロジーやステータス管理などが重くなる
- このVPI数の制限が意外と厳しい
  - VPI数が大きくなるとMSTを選択する必要があるが、それでも物理構成によってはVPI数制限を超えることがある
  - チャネルかつTrunk接続で大量のVLANを扱うケースで特に引っかかりやすい。この場合はTrunkで扱うVLANを必要なVLANのみにしてVPI数を削減するなどの必要がある
    - switchport trunk allowed vlan
  - VPI指標を超える設定も入るがそのうち処理できない状態となる可能性があることに注意

Sony Global Solutions

考慮点: スパニングツリー

- スパニングツリーはどうすれば…

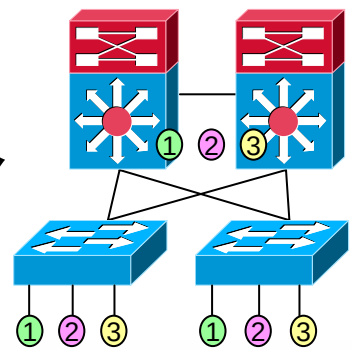


- IEEE802.1s (MST)などの多重化プロトコルを使う必要がある

考慮点: MST (1/2)

- MSTIの割り当て設計は十分に検討する
  - IEEE802.1sでは64MSTIまで
  - トポロジーチェンジはMSTI単位 → 影響範囲をよく考える
  - MSTIに対するVLANの割り当てを後から変更するのは大変な作業

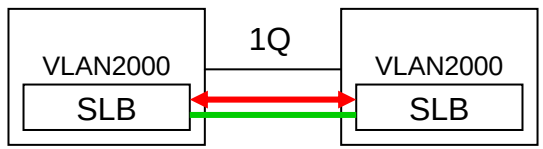
- 機器の仕様に注意
  - 規格準拠といいながらMSTI数を最大まで使えない実装もある
    → 16程度になっている場合がある
  - エッジポートの挙動などが規格と異なる場合がある



考慮点: MST (2/2)

- ハートビート用VLANに注意
  - ビルトインモジュールタイプのSLBなどを冗長構成で使う場合
  - ハートビートはVLANを通して行われる
  - MSTIを分けないとモジュールフェイルオーバーでMSTI内VLAN全体のトポロジーチェンジが起こる

  → ハートビートラインは物理的につなげたい

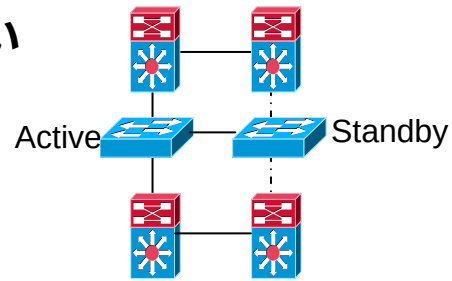


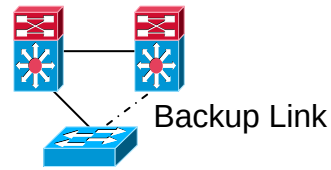
- MST以外のベンダー独自多重化プロトコルを使用する場合はその仕様に注意
  - インスタンス(ドメイン)に割り当てられる最大VLAN数
    - 合計で1000VLANまでのような制限がある場合がある
  - インスタンス(ドメイン)数増加によるコンバージェンス時間の増加

考慮点: スパニングツリー以外の方式

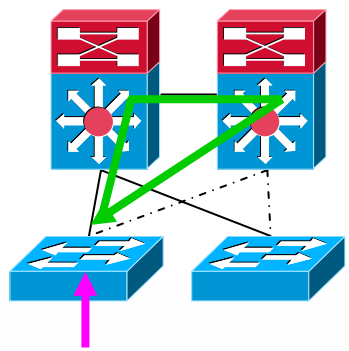
SONY

- (R)STP, MSTのトポロジーに参加{できない,したくない}場合
  - 機器がサポートしていない
  - MSTI内でさらにL2トポロジーを形成したい
  - 難しいことを考えたくない
- いくつかのベンダー独自方式がある
  - Active-Standby方式など
  - Cisco Flexlinkなど





- MACアドレステーブルの更新に注意
  - スパニングツリー以外ではMACアドレステーブルが更新されない場合がある
    - MACアドレステーブルのエージングアウトはデフォルトで300秒程度
      - エージングタイマーを短くするというのは…
  - エンドノードがルータのような機器であればISIS Helloなどでkeep-aliveするのもひとつのワークアラウンド
- ブロッキング処理方法に注意
  - 特に EtherType がIPv4以外のフレーム
    - IPv6
    - DECnet
    - …?



Sony Global Solutions

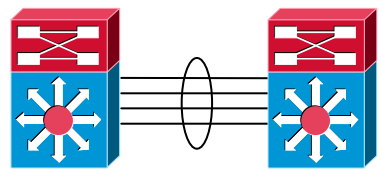
考慮点: リンクアグリゲーション

SONY

- スイッチ間のみならずスイッチ以外の機器やブレードサーバなどでも使用
- LACPなどの動的プロトコルを実装していない機器に注意
  - 通信不能になる場合がある
    - 論理I/Fが上がる前に物理I/Fにパケットが来るとブロードキャストしてしまう実装がある
    - 物理I/Fと論理I/Fを同時に上げる実装もある

    → MACアドレスのミスラーニング → 間違ったセッションテーブル作成
  - アグリゲーション外れがすぐにはわからない

→ リンクアグリゲーションを実用的に使用するには動的プロトコルが必須?



- その他
  - アグリゲーションを外していくとトポロジー内のリンクコストが減ってトポロジーが変わることがあるので特にメンテナンス時に注意
    - 帯域が変わるとリンクコストが変わる
    - PVST+とMSTでもまた違う

    ⇨ コスト値を明示的に設定しておくのが吉
  - パケットリオーダーに気をつける
    - パーパケットバランシング → 追い越しが発生することがある

Sony Global Solutions

考慮点: IPチェックサム

SONY

- L2なのになぜL3のチェックサムが問題になるのか?
  - CoSのマーキング時にIPチェックサムを再計算する実装がある
  - この実装によって相互接続に問題が出ることがある
- IPチェックサムの計算方法についてはいくつかのRFCで述べられている
  - RFC1071, 1141, 1624
  - アルゴリズムは1の補数和の1の補数

1の補数

-2	-1	-0	+0	1	2
0xFFFD	0xFFFE	0xFFFF	0x0000	0x0001	0x0002

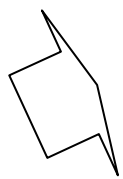
- チェックサムが+0になるときが問題
  - ヘッダ次第の場合と、特にSrc/Dst以下の情報が同じパケットを連続的に送出する場合
    - IPヘッダ内でSrc/Dstが同じでも識別番号フィールドはインクリメントされていく
      - 識別番号はデータグラムの送信順に1つずつ増加する
      - よってチェックサムも増加し、いずれは+0になる

    → Keep-aliveなどが該当しやすい
- チェックサムが+0のときに-0に付け替える実装がある
  - そして-0をドロップする実装がある
- チェックサム計算はASIC実装であることが多いので発覚したときには…

Sony Global Solutions

つまるところ

- 箱
  - キャパシティ
    - VLAN数は何も考えなくても最大値まで使えてほしい
    - MSTではなくPVST+でも使えてほしい
  - CPU処理
    - BPDUの送受信処理などもっとCPU処理部分を高速化してほしい
      - フラグメントやマルチキャストなどCPU処理部分は多い
  - 相互接続性
    - より一層の改善を…
- Ethernetそのもの
  - 誰かTTLを…
  - BPDU性善説というのもどうなのか?
  - IEEE802.1ah, ITU-T Y-17.ethoam?
  - …



とりあえず○○ガードなどを
ひととおり入れておくが…

いつまでSTPなのか…

最近気になっているもの

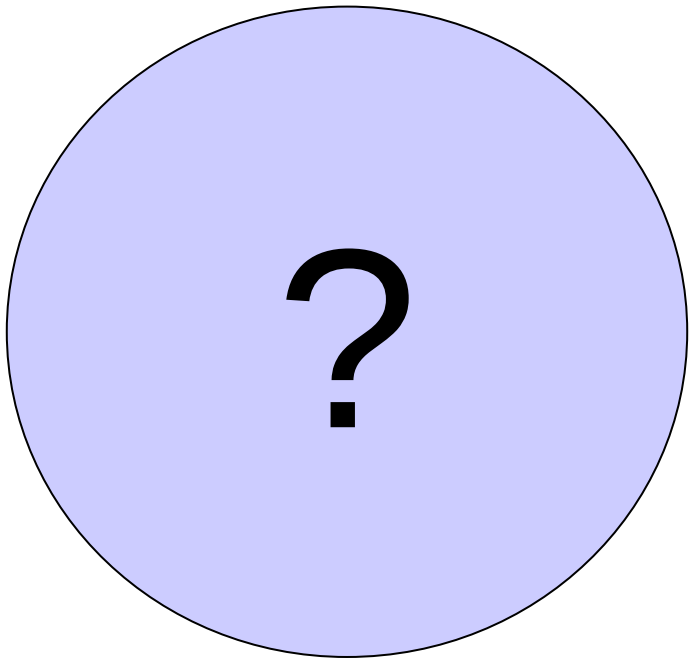
- TRILL (Transparent Interconnection of Lots of Links)
  - Motivation:
    - EthernetのIEEE802.1を改善したい

    → Ethernetでステートフルにトポロジー管理を行いたい
  - Architecture:
    - Forwarding based on safe header
      - TTL in the encapsulation header
    - Coexist with existing bridges
      - Outer header is Ethernet
    - Runs a link state routing protocol → IS-ISベース?
- GMPLS for Ethernet
  - 自由にパスが張れるようになるとおもしろい?

まとめ

- EthernetでLANを設計・構築・トラブルシュートするたびに憂鬱になる
  - 考えないといけないことが多い
  - 原因究明もたいへんなことが多い
  - トラブルを避けるために結局冗長化構成を見直すこともある
- Ethernetで大規模LANを設計・構築するのはまだまだ難しいと感じる
- また新しい技術も出てきているのでその動向に注目したい
- Ethernetはこれからどこへ向かうべきか?

Comments?



mailto:Toshio.Hiraga@jp.sony.com

Sony Global Solutions