SONY

コンバージェンスを重視したMPLSの美味しい使い方

ソニーグローバルソリューションズ株式会社
平賀十志男<*Toshio.Hiraga@jp.sony.com*>

JANOG14

Sony Global Solutions

概要

- ネットワークの可用性に対する要求は年々高まっており、障害発生時といえども高速に回復することが求められてきています。
- しかし、BGPなどの従来のルーティングプロトコルによるコンバージェンスはそれなりの時間がかかってしまうものです。
- 本発表ではMPLSの特徴を利用してコンバージェンス時間を重視した可用性の高いネットワークの実装例を紹介します。

会社概要

SONY

- **商号**: **ソニーグローバルソリューションズ株式会社**
- **設立**: 1988**年**2**月**
  - 2003**年**7**月合併に伴い、現社名となる**
- **資本金**: 3**億円(ソニー(株)**100%)
- **売上高**: **約**633**億円**(2003**年度**)
- **代表取締役社長**: **戸髙 修**
- **従業員数**: 1,029**名**(2004**年**7**月**1**日現在**)

Sony Global Solutions

ただし

- So-netではありません

SGS	AS9619	
SCN	AS2527	So-net
Sony	AS9600	bit-drive

- ソニーグループ国内250拠点の接続
- ソニーグループデータセンターオペレーション
  - 国内および北米
- ソニーグループ海外拠点の接続
  - 北米、欧州、アジア

飽くなきユーザ要求

SONY

- 30秒断で工場のラインや物流が止まる!?
  - 国内ネットワーク、国際ネットワーク
    - 地理的結びつきではなく、ビジネスの結びつき
  - アプリケーションのタイムアウトが短い
    - メインフレームの感覚
- 10秒断で問い合わせの電話
- パケットドロップは困る
  - Non IPプロトコルのIPカプセル化

Sony Global Solutions

IGPトラフィックエンジニアリング

SONY

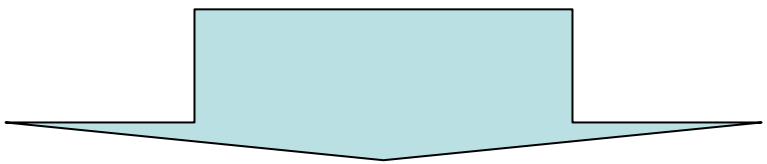
- モチベーション
  - サービスレベル、課金方式の異なる回線の使い分け
  - 通常時トラフィックのないリンクの存在
- やればある程度はできますよね。でも、…
  - たいへん?
    - 「この拠点はECMPにして、この拠点はメイン回線に寄せて…」
    - 「そのためにこの経路のコストはこの経路のコストより大きいのでここのコスト値を変えて…」
    - 「それだと、こことここが通信する経路のコストが変化してこの経路を通ってしまいます…」
    - 「では、全回線と全ルータのコストを表にしてECMPとなるようにしてみて…」
    - 「対象経路がどれだけあってルータが何台あるか知ってますか? もしかしていじめ?」

Sony Global Solutions

ネットワークの構築要件

SONY

- コンバージェンスをもっと速くしたい
  - 特にBGP
- ついでにトラフィックエンジニアリングもしたい



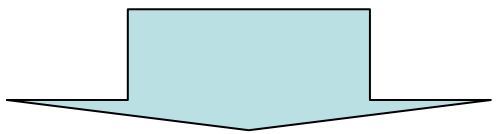
そうだ、MPLS

Sony Global Solutions

MPLSモチベーション

SONY

- **高速大容量ルータが欲しい**
- ATM**のコネクション指向と**IP**を統合したい**
- ATM**の**QoS/CoS**機能を**IP**で使いたい**



- Multi-Link Extension (IP over Everything)
- VPN
- **トラフィックエンジニアリング**
- **光波長**

Sony Global Solutions

MPLS-TE

SONY

- IGPの最短経路に依存しない明示的なルーティング
- 帯域や任意のリンク属性に基づいた最適パスによるリソースの有効活用
- 高速障害復旧機能

Sony Global Solutions

MPLS-TEの切り替え方式

SONY

- Global Repair
  - Restoration
    - **障害検知後、トポロジーの再計算、シグナリング、LSPの再確立**
  - Path Protection
    - **障害検知後、あらかじめ用意したバックアップLSPへ切り替え**
- Local Repair (FRR)
  - draft-ietf-mpls-rsvp-lsp-fastreroute-06.txt
  - One-to-One Backup
    - 1:1 Protection
  - Many-to-One (Facility) Backup
    - N:1 Protection
    - Link Protection
    - Node Protection

Sony Global Solutions

MPLS-TE実装例の概要

SONY

- Ethernetインターフェース
- ISIS-TE
- シグナリング
  - RSVP-TE w/ Fast Hello
  - 帯域指定などは特にしない
- 明示的なパス指定
  - 固定料金回線をメイン、従量課金回線をバックアップとするとか
  - 高品質回線をメイン、Not So高品質回線をバックアップとするなど
- 高速回復
  - リンクプロテクション
  - ノードプロテクション
- ユーザルート用ルーティングプロトコル
  - TE上ではeBGP, iBGPのみを使用
  - BGPルータをHead End/Tail Endとする

Sony Global Solutions

ISIS-TE?

- 感覚的に見ると
  - 格好よさそう?
  - 安定している?
    - 北米の大規模キャリア/ISPで使われているので安定している？
    - 以前のJANOGでNTTのフレックスギガウエイサービスもISIS-TEを使っているという話を聞いた
  - 実装が早い?
    - Optical Extensionは違ったけど
  - 最近北米メーカがISISへの移行を推進している気がする
- 技術的に見ると
  - デフォルトではOSPFより障害検知が速い
  - DR(DIS)をプリエンプトできる
  - データベースのLife Timeが20分で、しかも変更できる(1秒〜65535秒)
    - OSPFではデータベースのMaxAgeが1時間で、しかも変更できない
  - セキュリティ
    - OSPFはコントロールプレーンがIPベースなのでMD5認証が必要
    - ISISはコントロールプレーンがIPベースではないので到達性の点で有利?

BGP over MPLS-TE

SONY

- コンバージェンスポリシー
  - BGPルータではなるべくBGPセッションを落とさない
    - 回線ダウン、リンクダウン、途中ノードダウン時
      - MPLS-TEのプロテクション機能による高速切り替え
      - 特にフルルートを保持する場合は重要
  - 迂回できないなら高速に障害検知し、BGPセッションを落とす
- コストの高いBGPのコンバージェンスを極力使わず回復させることができる
- iBGPでもIGPの難しいことを考えずにトラフィックエンジニアリングができる

Sony Global Solutions

MPLS-TE上のeBGP

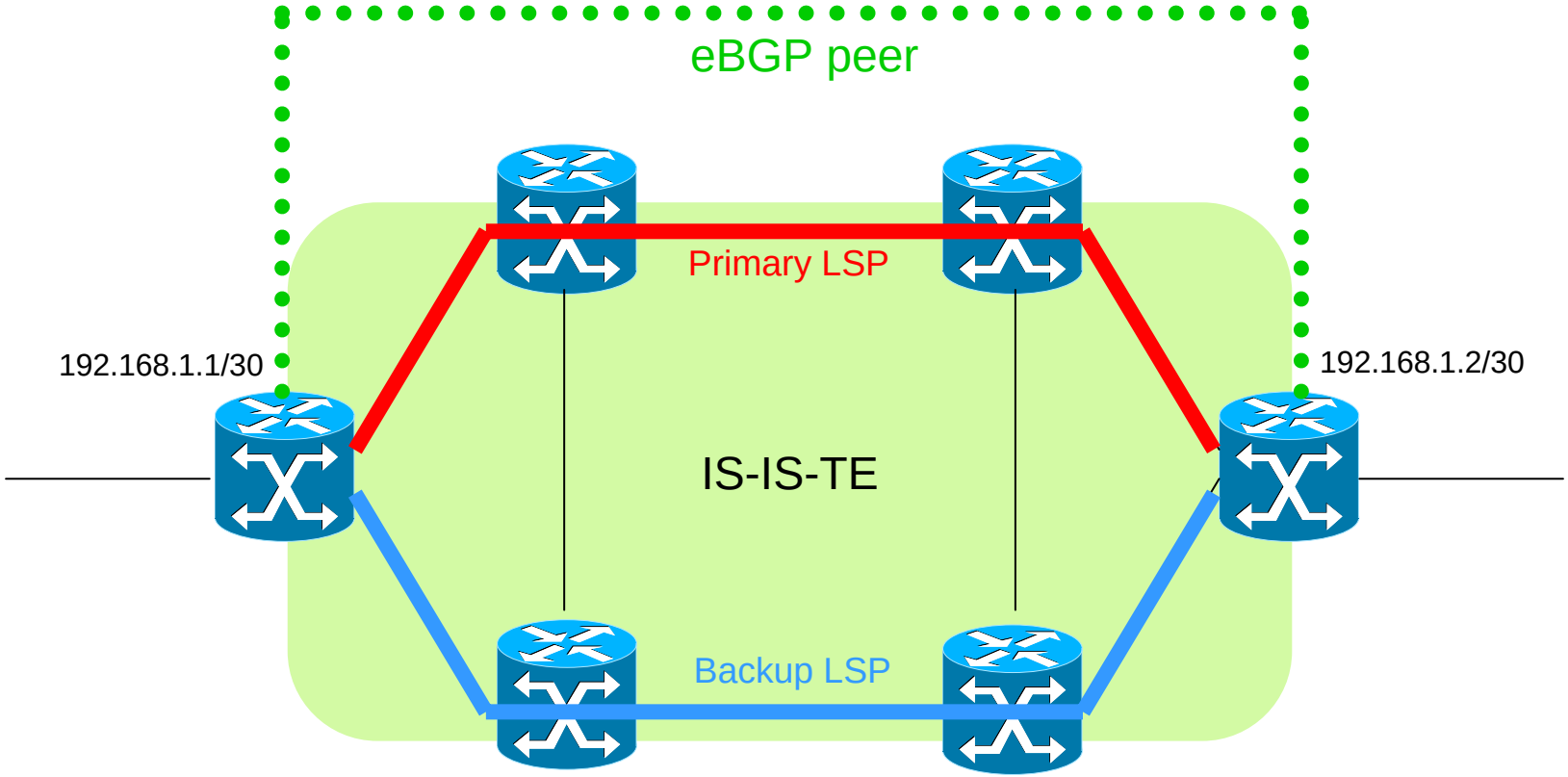
SONY

- Numbered TE Tunnelを使う
  - /30
  - Tunnelインターフェースダウンで BGPが障害を検知できる
    - C3o社の実装ではTEトンネルもI/Fに見える
    - BGPホールドタイマーを待ちたくない(<180s)

      fast external fallover
  - スイッチのような機器が間に入っていてもリンク障害を検知できる

Sony Global Solutions

eBGP概念図



設定例(eBGP) C3o社の場合

SONY

```
interface Tunnel0
 ip address 192.168.1.1 255.255.255.252
 tunnel destination 10.1.1.2
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng path-option 1 explicit name ebgpprimary
 tunnel mpls traffic-eng path-option 2 explicit name ebgpsecondary
 tunnel mpls traffic-eng path-option 3 dynamic
 tunnel mpls traffic-eng record-route
 tunnel mpls traffic-eng fast-reroute
!
interface Tunnel1
 ip unnumbered Loopback0
 tunnel destination 10.1.1.2
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng path-option 1 explicit name ebgpsecondary
 tunnel mpls traffic-eng record-route
!
router bgp 65000
 neighbor 192.168.1.2 remote-as 65001
 neighbor 192.168.1.2 password 7 ABCDEFGHIJKLMN
 neighbor 192.168.1.2 timers 30 90
 neighbor 192.168.1.2 advertisement-interval 1
!
mpls static binding ipv4 192.168.1.0 255.255.255.252 10002
```


Sony Global Solutions

Numbered Tunnel使用時の注意点

SONY

- **現状、**numbered tunnel**とLDPを併用する場合にはあらかじめラベルを予約しておかなければならない**
  - EoMPLS(Martini)**との併用など**
- Tunnel**インターフェースがすぐには落ちない**!?
  - RSVP**では障害検知しているにもかかわらずなかなか落ちない**(30**秒弱**)
  - Tunnel**インターフェースのダウンに謎のタイマーがある**?
    - MPLS**に限らず**Tunnel**インターフェースを使うものすべて**
    - **しかも固定値**?

落ちたのなら早く落ちてほしい

Sony Global Solutions

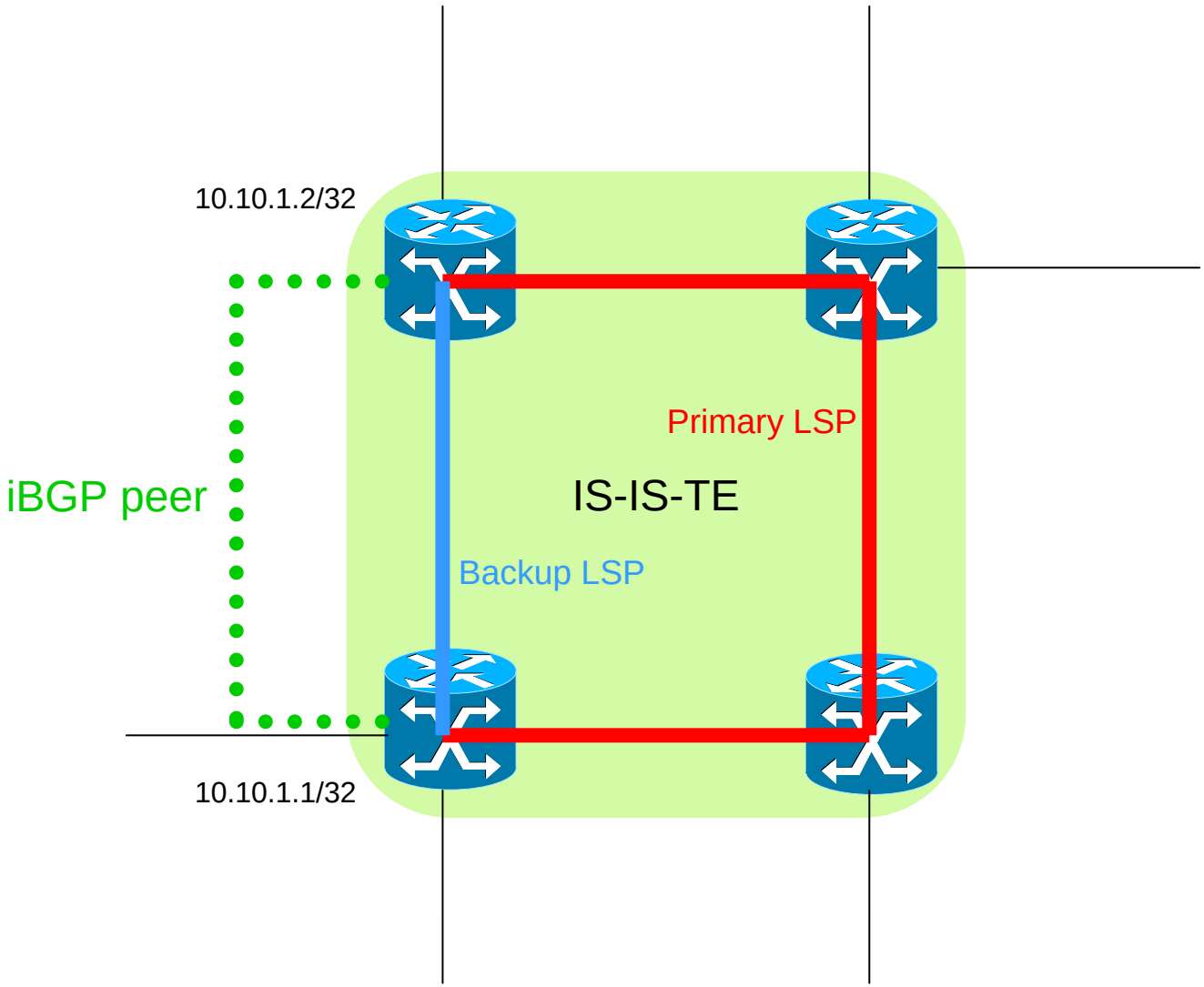
MPLS-TE上のiBGP

SONY

- Unnumbered TE Tunnelを使う
  - ループバックアドレスでBGP接続を行う
  - BGPでTEトンネルを使うように静的経路を使う
    - auto-route announceはBGP以外もTEトンネルを通るため、コントロールプレーンのトラブルシューティングを考えて静的経路とする
  - IGPでの障害通知およびSPFの再計算がFRRに必要ないためIGPよりもコンバージェンスが速い
  - それでもBGPスキャンタイマーがあるので、BGPネイバーダウンには効果がない
    - next-hopが無効になれば即座にBGPに通知してほしい

Sony Global Solutions

iBGP概念図



設定例(iBGP) C3o社の場合

SONY

```
interface Tunnel100
 ip unnumbered Loopback0
 tunnel destination 10.10.1.2
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng path-option 1 explicit name ibgpprimary
 tunnel mpls traffic-eng path-option 2 explicit name ibgpsecondary
 tunnel mpls traffic-eng path-option 3 dynamic
 tunnel mpls traffic-eng record-route
 tunnel mpls traffic-eng fast-reroute

interface Tunnel101
 ip unnumbered Loopback0
 tunnel destination 10.10.1.2
 tunnel mode mpls traffic-eng
 tunnel mpls traffic-eng path-option 1 explicit name ibgpsecondary
 tunnel mpls traffic-eng record-route
!
router bgp 65000
 neighbor 10.10.1.2 remote-as 65000
 neighbor 10.10.1.2 password 7 ABCDEFGHIJKLMN
 neighbor 10.10.1.2 update-source Tunnel100
 neighbor 10.10.1.2 next-hop-self
 neighbor 10.10.1.2 timers 30 90
 neighbor 10.10.1.2 advertisement-interval 1
!
ip route 10.10.1.2 255.255.255.255 Tunnel100
```


Sony Global Solutions

FRRは高速?

- POSとEthernetでは回復時間が異なる
  - 現在はRSVP Fast Helloが一番早く検出
  - Ethernetでもさすがに1秒はかからなくできるが…
- Ethernetでも改善したい
  - キャリア検出時間を短くするのは?
    - 0秒にしても検出に時間がかかる(xxx ms?)
    - リンクフラップのときに困る
      - C3o社の実装ではdownとupで検出時間を変えられないので、そもそも0秒にはしにくい
    - FRRとルーティングコンバージェンスは分離したい
      - リソースをFRRに集中させるほうが安全
  - RSVP Fast Helloの間隔を短くするのは?
    - ラインカードでオフロードできるようにならないと厳しい?
  - EthernetでもLoS等でFRRが作動してほしい
- これ以上を望むならやっぱりPOSですか?
  - もちろんコスト大幅増…

ところでJ5rだと?

- RSVP Fast Helloがない
  - L2アラーム
  - ノードダウン検知はほかの仕組みで
    - IGP Update
    - RSVP Path Refresh Timeout (最小1秒×3)
  - MPLS BFD待ち?
- Ethernetのリンクダウン
  - デフォルトの0秒でも検出に時間がかかる(xxx ms?)
  - キャリア検出はdownとupで時間を変えられる
    - リンクフラップ対策には有利
- そもそもTEトンネルはインターフェースではない
  - /30はできない
  - TEトンネルインターフェースダウントリガーではなく、TEトンネルとnext-hopの連動によってピア先のループバックアドレスがinvalid next-hopとなり、ホールドタイマーの時間切れを待たずに即BGPコンバージェンスができる
- BGPスキャンタイマーの概念がない?
  - IGPでもInvalid next-hopで即BGPコンバージェンス

微妙

RSVPによるTEの障害検知はC3oより遅いが、TEの障害を検知してからBGPのコンバージェンスが動くまではC3oより速い

MPLS-TE構築上の注意点

SONY

- **回線の最大フレームサイズに気をつける**
  - 「1522バイトでよかったですよね?」

DA(6)+SA(6)+Type(2)+Label(4)+Payload(1500)+FCS(4)=1522

通し!

しかし、FRRすると

DA(6)+SA(6)+Type(2)+FRR Label(4)+Label(4)+Payload(1500)+FCS(4)=1526

ロン!

Sony Global Solutions

トラブルシューティングのために

SONY

- これくらいのログは取得しておくと便利

```
mpls traffic-eng logging lsp path-errors
mpls traffic-eng logging lsp reservation-errors
mpls traffic-eng logging lsp preemption
mpls traffic-eng logging lsp setups
mpls traffic-eng logging lsp teardowns
mpls traffic-eng logging tunnel lsp-selection
mpls traffic-eng logging tunnel path change
```


Sony Global Solutions

まとめ

- コンバージェンスという観点でMPLS-TEの活用方法とその実装例を示しました
- IPルーティングだけでがんばるよりも、プロテクトしたいところだけプロテクトして高速に切り替えるというのが美味しいと思っています
- BGPセッションがそう簡単には落ちない分、可用性は高まるのではないかと思います
- MPLSのことがあまりわからなくてもとりあえず使えます
- MPLS-TEの活用例を示しましたが、ほかに面白い使い方があったら教えてください

おしまい

- ありがとうございました
- ご意見、ご質問、間違い等がありましたらお願いします