

運用から見た研究 研究から見た運用

日本電気株式会社 クラウドシステム研究所
金子 紘也

自己紹介

名前

金子 紘也 @社会人二年目

所属

日本電気株式会社 クラウドシステム研究所

JANOG歴

本会議初参加 (Interim含めると2回目)

業務

広域ネットワークへのSDN適用に関する研究開発

- 最近は最適パスの構成アルゴリズムなど

本セッションの概要

「深く尖った」フィールドである研究開発にフォーカス

■ 運用と研究を取り巻く環境の変化について(10分)

- ネットワーキング分野における研究の概要
- HyperGiantにおける運用と研究のサイクルの一体化

■ 国際学会から、運用を変えうる研究を紹介(20分)

- ISP / WAN / DC / IX と対象領域別に4つ紹介
- 研究者から見た、運用を変えうるCoolな研究をチョイス

■ みんなでディスカッション(30分)

1. 運用者から見た研究事例への率直な意見
2. 運用と研究のサイクルを回すためになにをすべき？

ネットワーク分野における研究

国際学会、論文誌などで研究発表

計算機系研究は国際学会も論文誌同様重視される

ACM SIGCOMM 2014

- Data Plane
- Network Architecture
- Middleboxes And Network Services
- Wireless
- Monitoring And Diagnostics
- Novel Datacenter Network Designs
- Scheduling In Datacenter Networks
- Network Operations
- Transport and CC

その他 IEEE系だとGlobecom , ICC 等

USENIX NSDI 2014

- Datacenter Networks
- Debugging Complex Systems
- Software Verification and Testing
- Security and Privacy
- Operational Systems Track
- Data Storage and Analytics
- Interpreting Signals
- Improving Throughput and Latency
(at Different Layers)
- In-Memory Computing and Caching
- Scalable Networking
- New Programming Abstractions

(昔ながらの)運用の現場と研究の現場

運用の現場		研究の現場
実NW運用者	Player	大学、企業のR&D
運用の改善 / RFC	Contribution	学会発表 Journal Paper
運用の改善 / 相互運用性の確保	Mission	新規性 新技術の芽を出す
IETF / *NOG	field	SIGCOMM / NSDI

目的意識が違っていた

(昔ながらの)運用の現場と研究の現場

運用の現場

研究の現場

運用のこと考えて
なさそう...

そこはそんなに
困ってないし...

学会とかチェックし
てる時間ないし...

結局どれが主流に
なるの？

Player

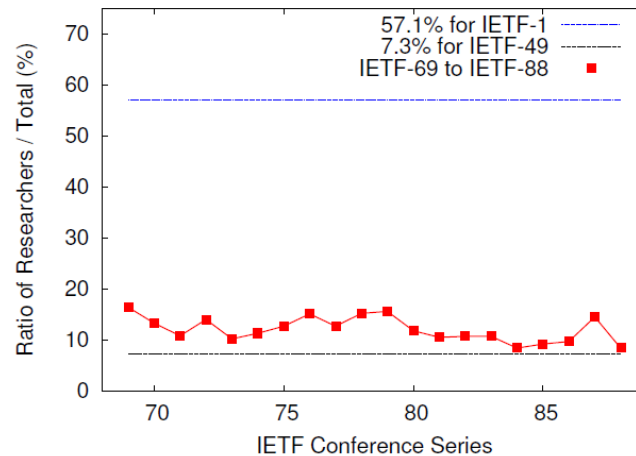


Figure 1: Researcher Participation at the IETF

人的な交流の少なさ

すごく良い技術(だと思ってる)
のに広まっていけない

Journal

実データ無いし...

とりあえず重箱の隅を
つついて業績を作ろう

昔はこれでもよかった(かもしれない)

研究の舞台における変化

国際学会におけるプレイヤーが変わりつつある

採択論文著者

- **NW系名門大学 + Microsoft + Google + Cisco**
- **Facebook, akamaiなどもコンスタントに発表**

Program Committee

- **NSDI/SIGCOMM 共に機器ベンダの存在感はほぼなし**

スポンサーシップ

- **機器ベンダのスポンサーシップは横ばいから減少傾向**
- **近年では、Facebook, akamai等が目立つ**

**サービス事業者が大きな存在感を
発揮する時代になりつつある**

HyperGiantにおける研究と運用の一体化

既に実運用している技術を報告する場としての学会

- 世界規模で既に動いているという説得力の高さ

実運用しているサービスを起点に**研究->開発->運用のサイクル**が一社の中で回っている

- 結果として、運用性を考慮した動く研究開発が行われる

必要なスイッチを自作して、
2011年から段階的にSDN化した広域網を実運用 @Google

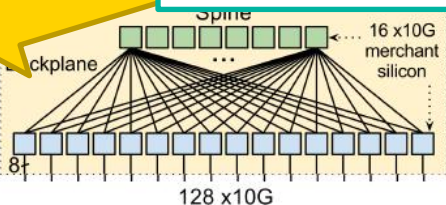


Figure 3: A custom-built switch and its topology.



Figure 1: B4 worldwide deployment (2011).

日本は置いて行かれている??

本セッションの目的

国際学会における最新研究を紹介することで、

■ 運用者にとって

最新の研究事例紹介から、
運用を変えるかもしれない先端技術を知る。

■ 研究者にとって

運用者から見た研究を知ること、
実運用に向けて何が足りないのかを知る場に

■ 日本のインターネットにとって

研究->開発->運用のサイクルを回すきっかけに

紹介する研究

■ 研究者目線で「運用を変えうる研究」をPickup

- 最近のSIGCOMM/NSDI等から紹介
- 研究者目線でCoolだと思った研究

■ 4つの適用分野ごとに紹介

- DC - データセンタに閉じたNW運用技術など
- Internet/ISP - AS間接続を含むNW運用技術

金子が担当

- WAN - 広域網(DC間接続)におけるTEなど
- IX - Internet Exchangeにおける運用技術

石田が担当

紹介する研究

■ 研究者目線で「運用を変えうる研究」をPickup

- 最近のSIGCOMM/NSDI等から紹介
- 研究者目線でCoolだと思った研究

■ 4つの適用分野ごとに紹介

- DC - データセンタに閉じたNW運用技術など
- Internet/ISP - AS間接続を含むNW運用技術

金子が担当

- WAN - 広域網(DC間接続)におけるTEなど
- IX - Internet Exchangeにおける運用技術

石田が担当

データセンタ領域

Statesman: A network-state management service

- Peng Sun et al. @SIGCOMM2014 Conference
- 著者構成: Princeton 2名、Microsoft 4名

目的

- ネットワーク運用タスクの安全な自動化
 - スイッチのFirmware updateの自動化などが可能に

手法

- 既存のプロトコル(BGP)が動作するNWのモデル化
- モデル上での依存関係の自動解決

結果

- Azure DCで既に10ヶ月稼働中(学会発表時点)
 - 10 DCs 20K devices

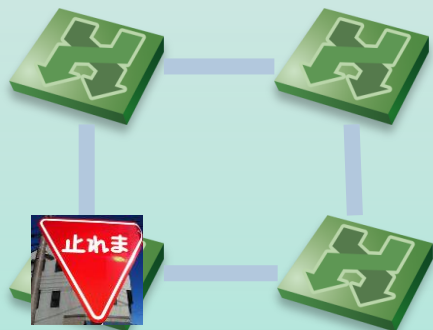
NW運用の自動化

■ 本研究で大きな対象にしているのは運用タスク

- サービスプロビジョニング系だけに限らない

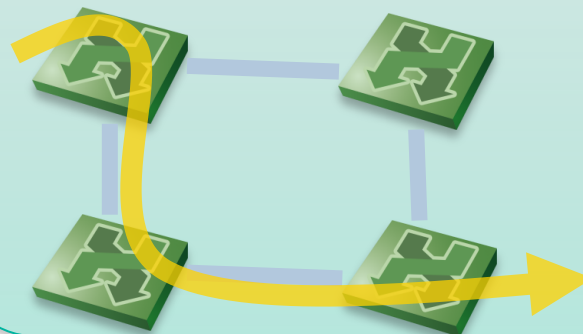
Firmware upgrade

- 古いFWのスイッチをアップデート



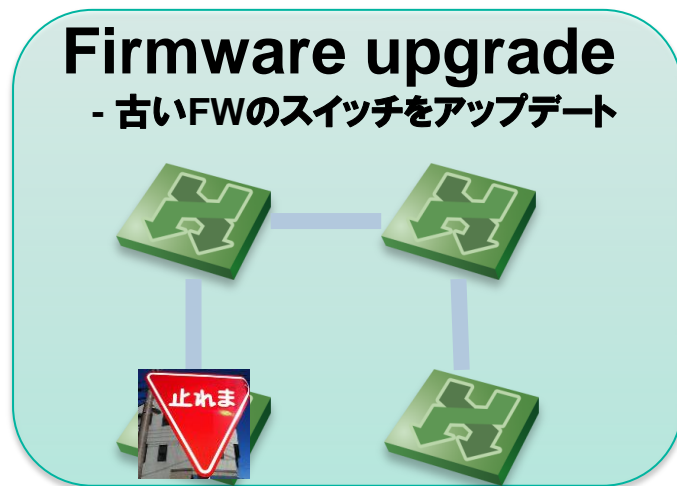
Traffic engineering

- demandからパスを最適に変更



運用タスク自動化の課題1

異常系への対応



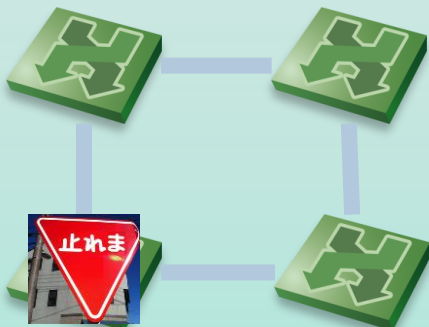
**ファームアップ後再起動を繰り返したら？
ファームのダウンロードに失敗したら？**

運用タスク自動化の課題2

タスク間のコンフリクト問題

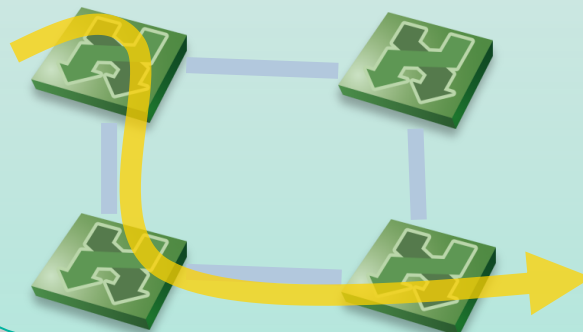
Firmware upgrade

- 古いFWのスイッチをアップデート

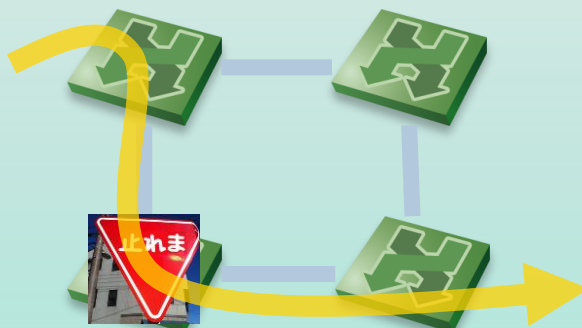


Traffic engineering

- demandからパスを最適に変更



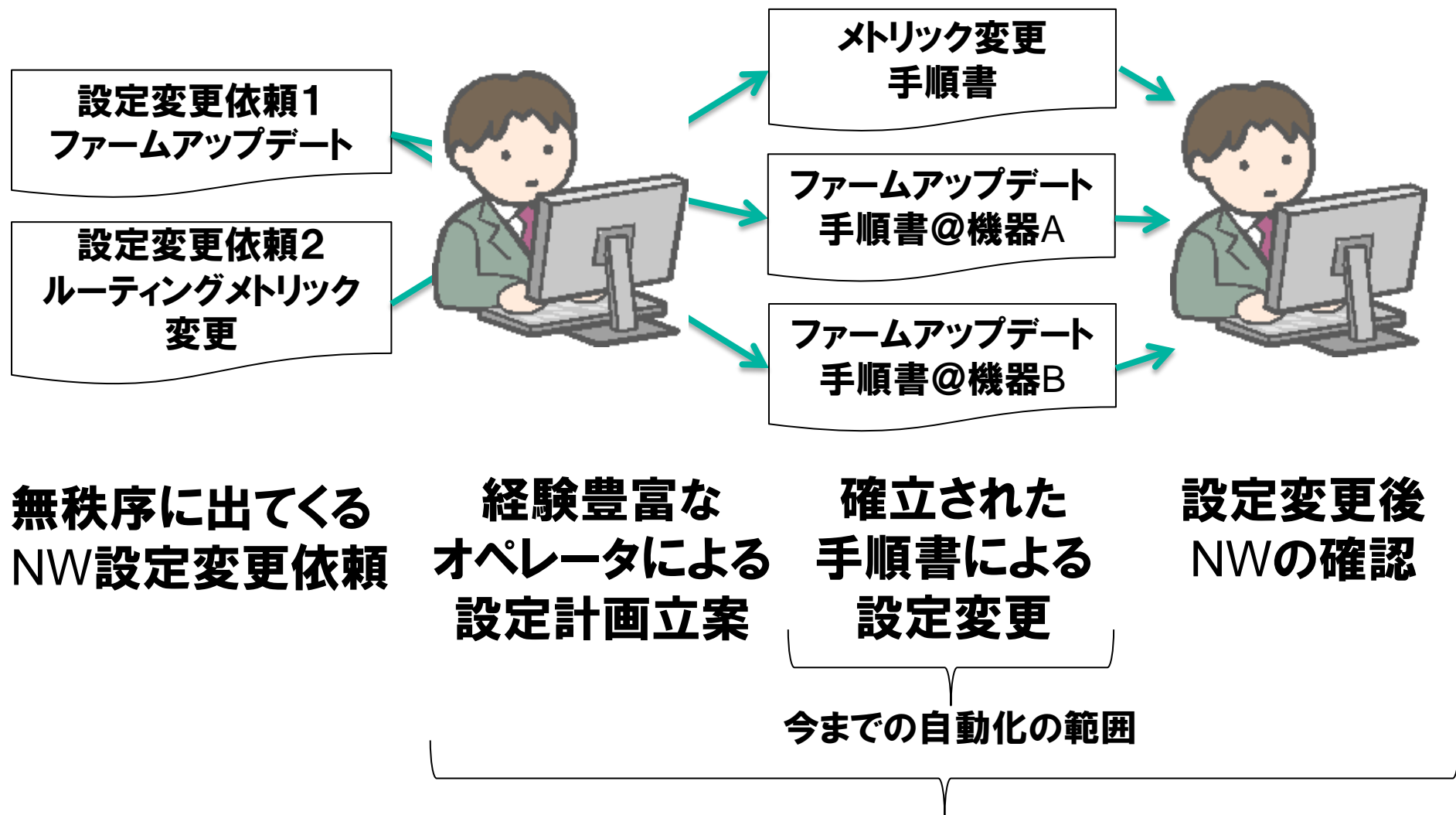
物理NW



これじゃ困る

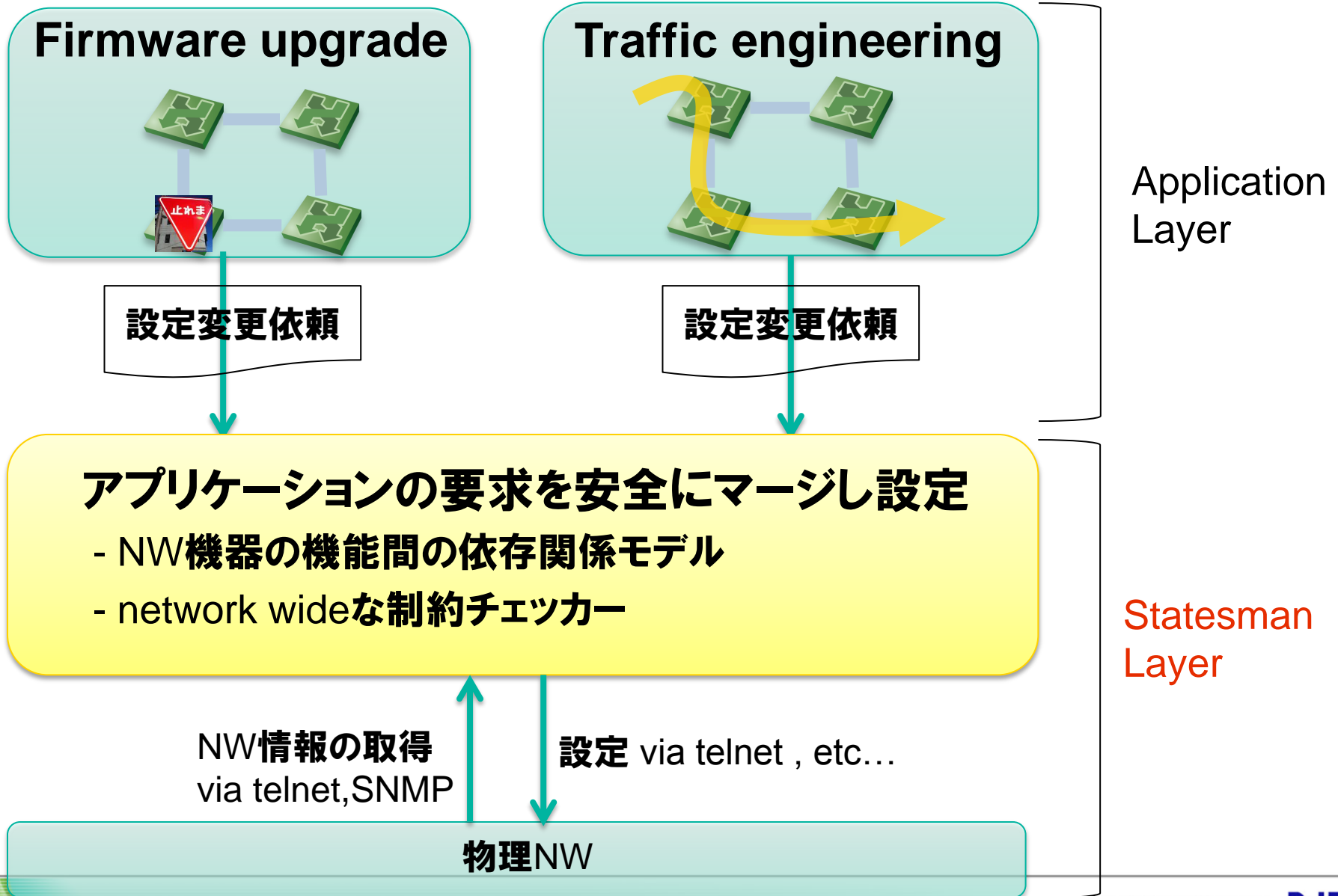
やっぱり手動じゃないと危なそう？

手動だとなぜうまく行くのか



Statesmanが提供する自動化

Statesman Architecture



安全な実行計画の導出

装置単体の保証

NW機器機能の依存関係モデリング

電源入っていない装置の
メトリック変更できない

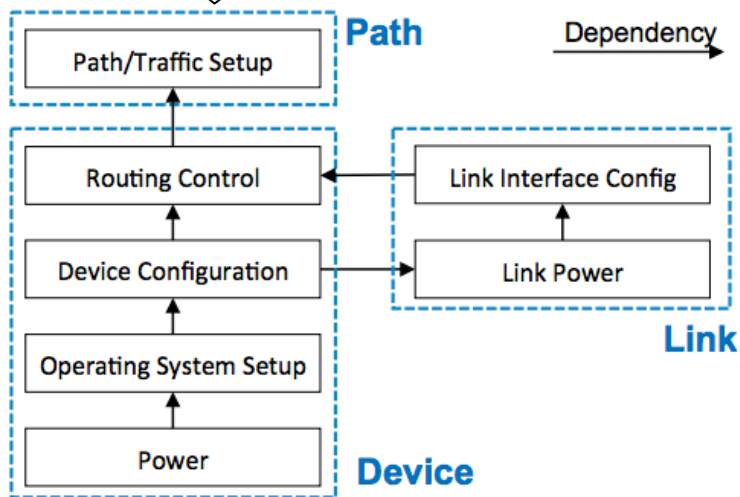
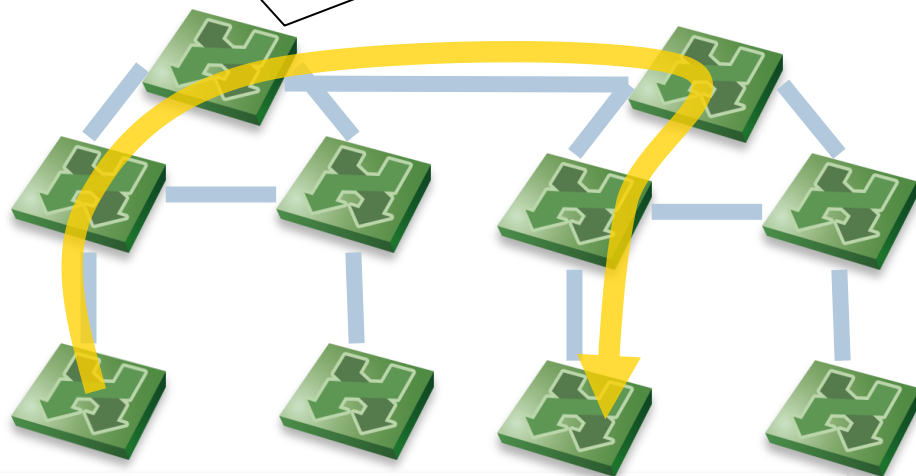


Figure 4: Network state dependency model

ネットワークワイドな保証

- ToR間のconnectivity, 帯域
- DC間のreachability

ファームアップ中にも
ToR間は性能を最低限保持して欲しい



可用性を維持した上で、最大限効率のよい
タスク実行計画を自動で立案

動作例

AzureのDCを跨る250スイッチのファームアップデート

Operatorの操作ミスによる
ファクトリーリセット

再起動後の不安定な動作

Full-Memoryによる
Firm Downloadの失敗

“自動化が苦手としていた”
と言われていた異常系

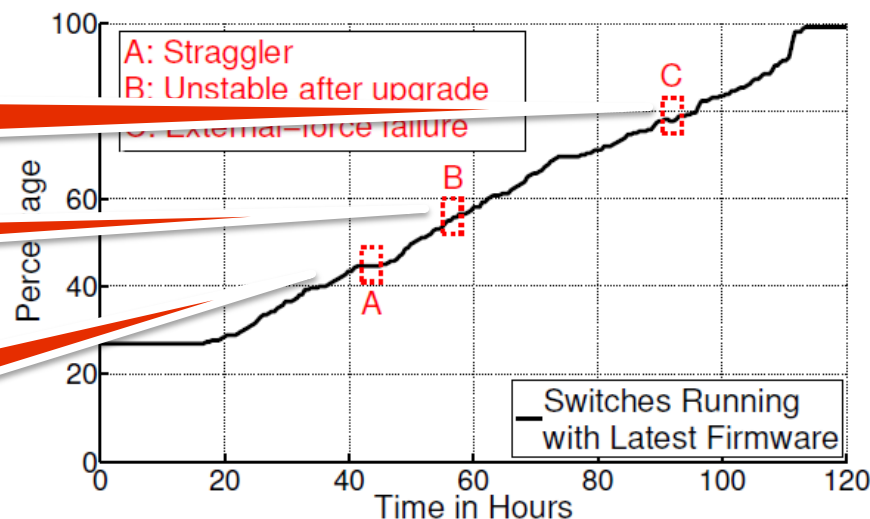


Figure 11: Time series of firmware upgrade at scale

250スイッチのアップデートを
120hで全て完了

ネットワークの可用性を保証しつつ、
自動的にFW Updateをスケジューリング

データセンタ領域まとめ

運用自動化の課題

- 異常系への弱さ，運用タスク間の依存関係

Statesman

- 複数のネットワーク運用タスクを安全に実行するための基盤
- タスクのモデル化と、タスク間のマージ、スケジューリング
 - ・ 機器の依存関係モデル，ネットワークワイドな制約

ポイント

- **既存のNW機器に適用可能なSDN的アプローチ**
- オペレータによる手動適用に匹敵する**安全かつ高速なスケジューリング**を実現
- **実際にAzure DCで動作しているという実績**

紹介する研究

■ 研究者目線で「運用を変えうる研究」をPickup

- 最近のSIGCOMM/NSDI等から紹介
- 研究者目線でCoolだと思った研究

■ 4つの適用分野ごとに紹介

- DC - データセンタに閉じたNW運用技術など
- Internet/ISP - AS間接続を含むNW運用技術

金子が担当

- WAN - 広域網(DC間接続)におけるTEなど
- IX - Internet Exchangeにおける運用技術

石田が担当

Internet/ISP領域

Revisiting routing control platforms with the eyes and muscles of software-defined networking

- Christian Esteve Rothenberg et al. @SIGCOMM2012
- **著者構成**: CPqD 3名、UniRio 2名、NTT MCL 1名
- Routeflowと呼ばれていたシステムの改良

目的

- **エッジルータレスでAS構成**
 - 自ドメイン内経路制御の柔軟化、CAPEX削減

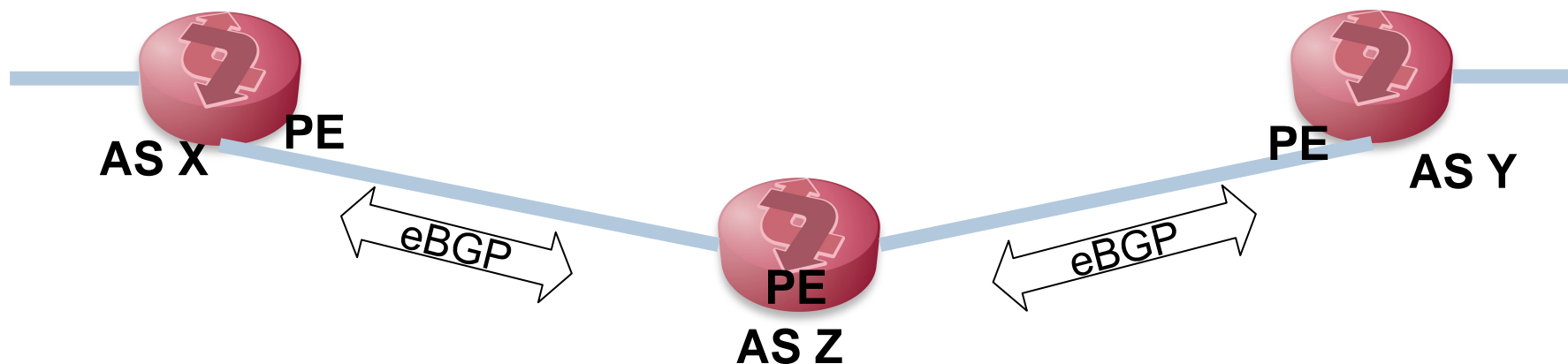
手法

- BGPのC/D-Planeを分離し、C-Planeをサーバにオフロード

ポイント

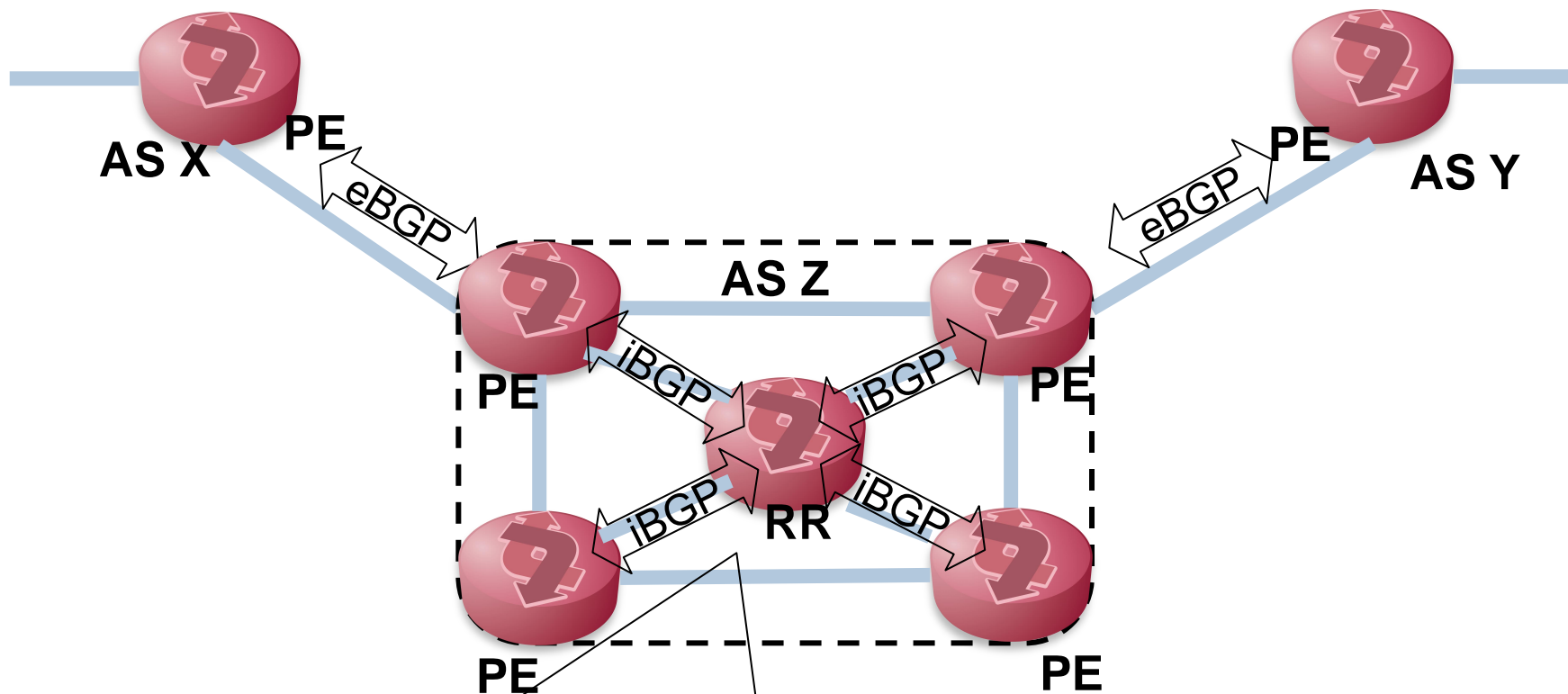
- SDNと既存システムとのいいところ取りをしたアーキテクチャ

インターネットの世界



Path Vectorベースの
シンプルでスケールするアーキテクチャ

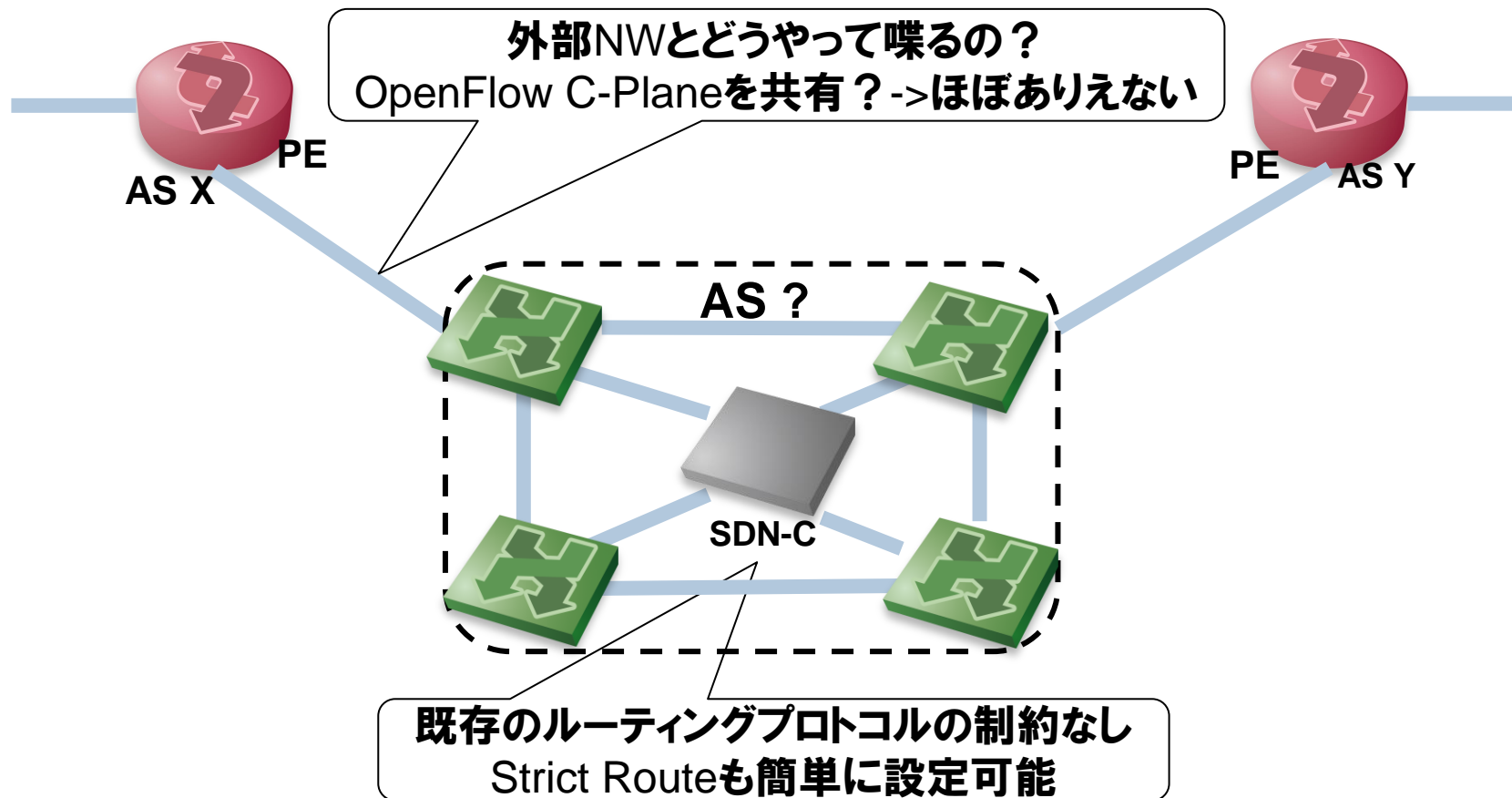
ドメイン内ルーティングに目を向けると



RouteReflection, BGP confederation...
IP/MPLSの場合LDPとかRSVPとかPCEPとか...
IP-VPNしたい場合はVRFで区切って...

既存のドメイン内経路制御は複雑

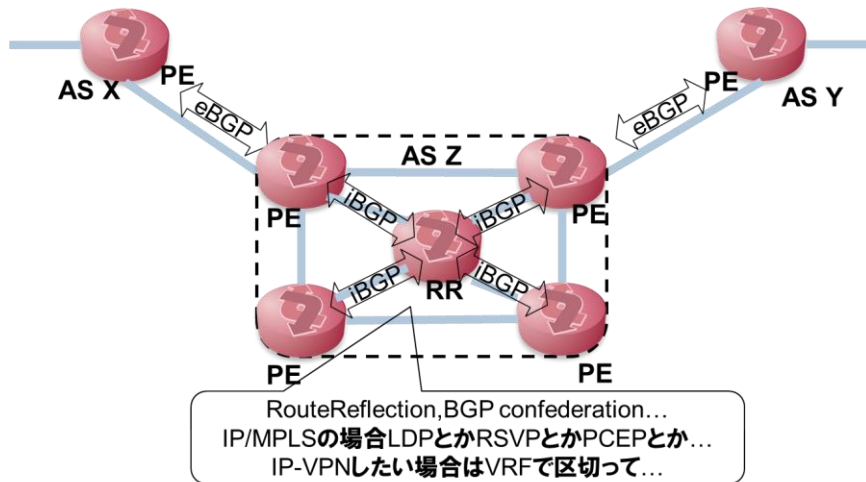
SDNの世界



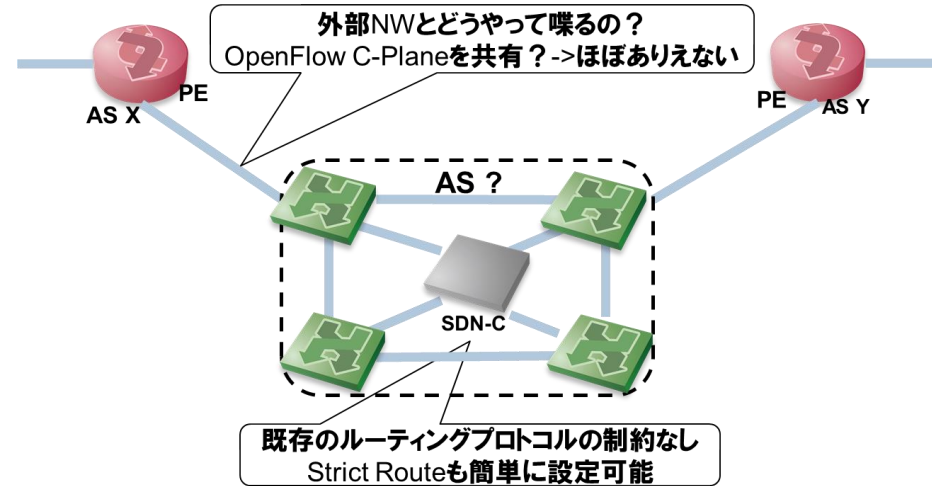
自ドメインは自由自在
一方、インターネット全体がSDN化する事はなさそう

得意な分野がそれぞれ存在する

BGP Based



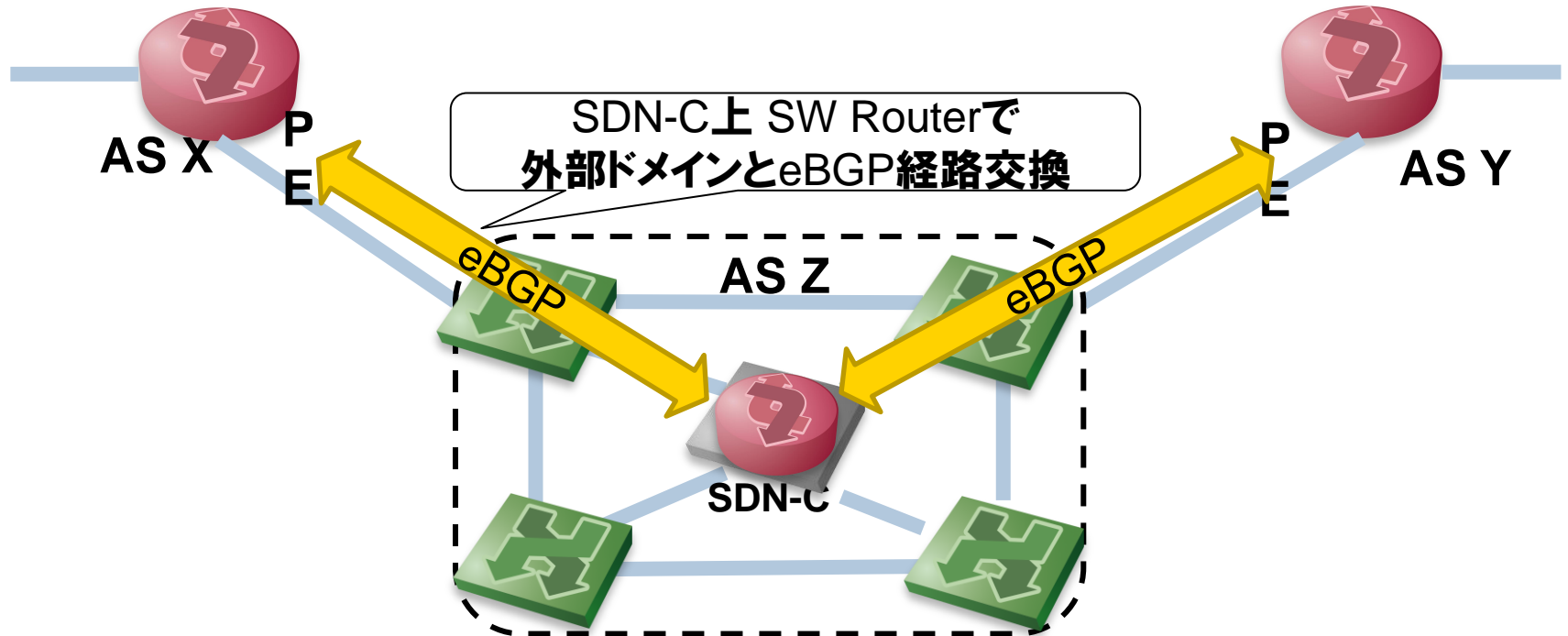
SDN Based



	ドメイン間	ドメイン内
BGP Based	○	△
SDN Based	×	○

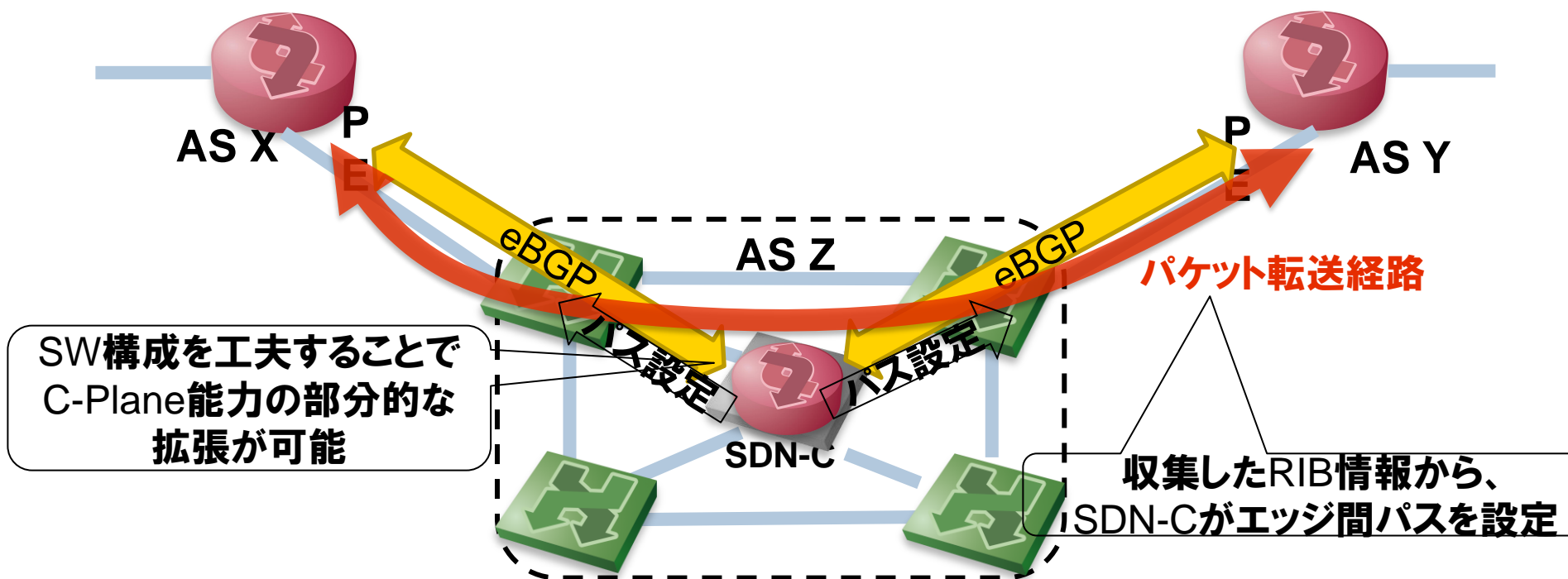
Hybrid Networking Model

1. 外部ドメインとの経路交換はeBGPで実施(C-Plane)



Hybrid Networking Model

1. 外部ドメインとの経路交換はeBGPで実施(C-Plane)
2. 内部ドメインはRIB情報から、コントローラが集中制御で転送パスを設定(D-Plane)



エッジルータ無しにASを構成

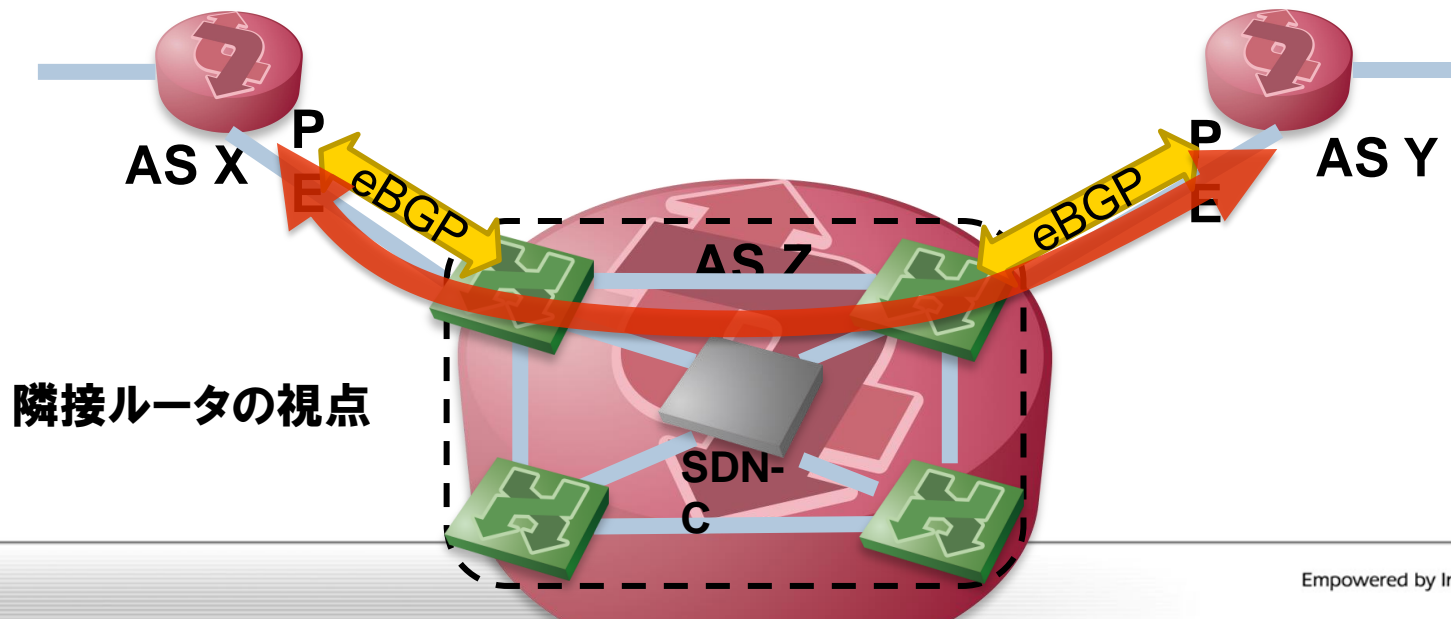
Hybrid Networking Modelの特徴

サーバ/スイッチがそれぞれ得意とする分野を担当

- プロトコル処理(C-Plane)は処理の柔軟性が高いサーバ
- データ転送(D-Plane)は高速なハードウェアで行う

その結果

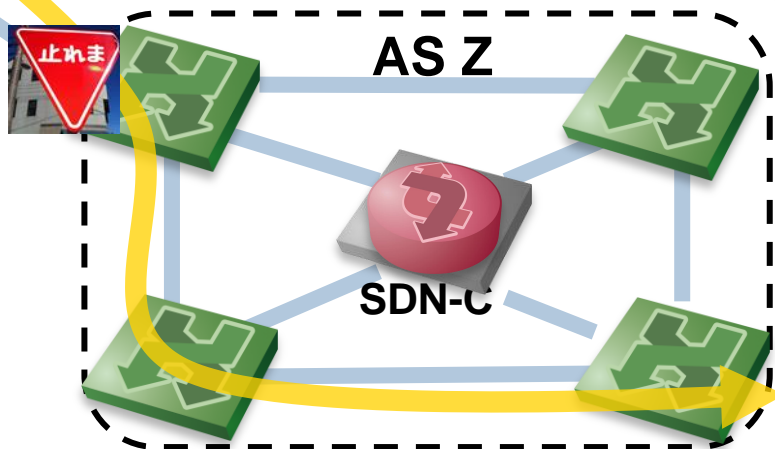
- 転送性能の劣化無しにスイッチ+サーバでIP転送を実現
- 外側からは既存のルータそのものに見える



ユースケース

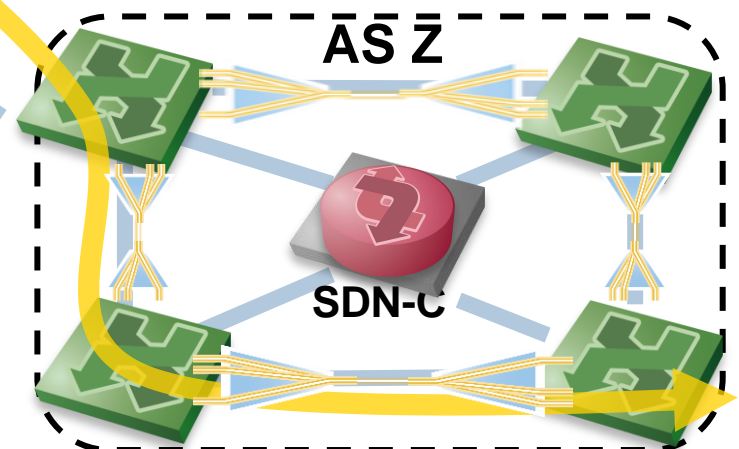
DDoS Mitigation

- より入り口に近いところで
- 精度よくdrop



E2E Directパス

- DWDM
- 伝送系をunderlayとして利用



ISP/Internet領域まとめ

Hybrid Networking Model

- **プロトコル処理(C-Plane)は処理の柔軟性が高いサーバで行い、**
- **データ転送(D-Plane)は高速なハードウェアで行う**
-> **エッジルータレスでのIP Forwardingを実現**

得られるメリット

- **装置コストの削減**
- **内部経路の柔軟な制御**
- **C/D-Planeそれぞれ分離した性能拡張**

ポイント

eBGPで接続する相手からはいままで通りルータがいるように見せつつ、内部ドメインの制御のみSDN化することが可能

Orchestrating a brighter world

世界の想いを、未来へつなげる。

未来に向かい、人が生きる、豊かに生きるために欠かせないもの。
それは「安全」「安心」「効率」「公平」という価値が実現された社会です。

NECは、ネットワーク技術とコンピューティング技術をあわせ持つ
類のないインテグレーターとしてリーダーシップを発揮し、
卓越した技術とさまざまな知見やアイデアを融合することで、
世界の国々や地域の人々と協奏しながら、
明るく希望に満ちた暮らしと社会を実現し、未来につなげていきます。

Empowered by Innovation

NEC