

SDN の段階的導入における 実装モデルの提案と、問題点の解決手法 【AP-GWモデルの提案】

中山 裕貴[†]，森 達郎^{††}，上野 哲^{††}

渡邊 儀秀^{††}，林 經正[†]

[†] 株式会社ボスコ・テクノロジーズ

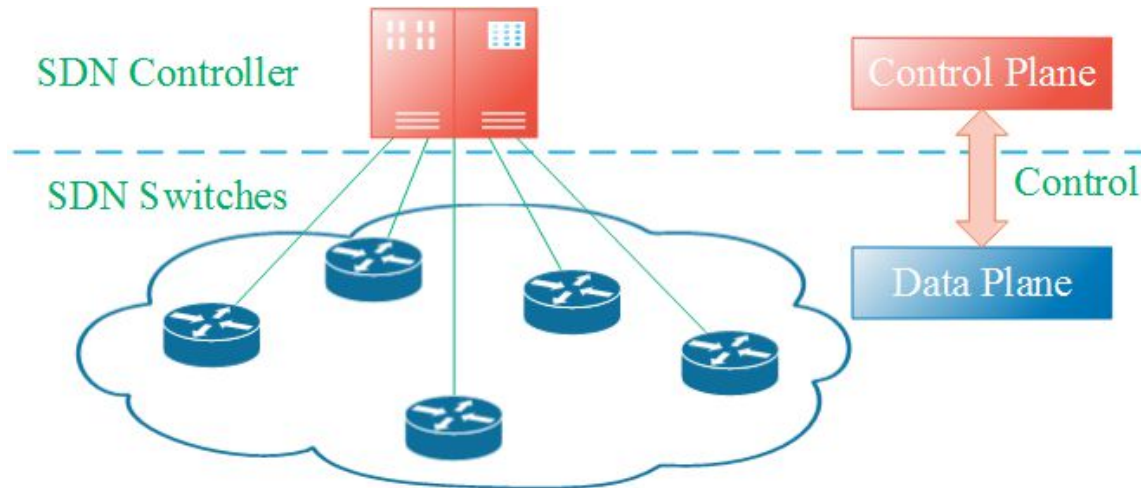
^{††} NTTコミュニケーションズ株式会社



研究背景

- Software-Defined Network (SDN) が注目されている
 - ネットワークの状況に応じてフロー単位の制御が可能
 - 従来からのネットワーク管理における問題解決への期待
 - ベンダ依存のスイッチの設定方法の違いによる複雑性の解消
 - 人手の介入が減少することによるヒューマンエラーの回避
 - ネットワークリソースの集約効率化と、エンドユーザ要望の迅速対応
- 実際に検討される応用例
 - データセンタ
 - ネットワークが即座に構築可能
 - ネットワーク帯域資源の有効利用
 - 広域網
 - サービス主導のネットワーク柔軟かつ、迅速制御
 - セキュリティ対策
 - DDoS攻撃への対応

現状のSDN実装モデルの問題点



- スケーラビリティの問題
 - 制御するネットワークの規模に応じてSDNコントローラへの負荷が増大
- 既存IPルーティングプロトコルへの対応
 - 閉じたネットワークでの実装
 - SDN導入のために全スイッチを総入れ替え
 - 対応することでSDNコントローラへの負荷がさらに増大

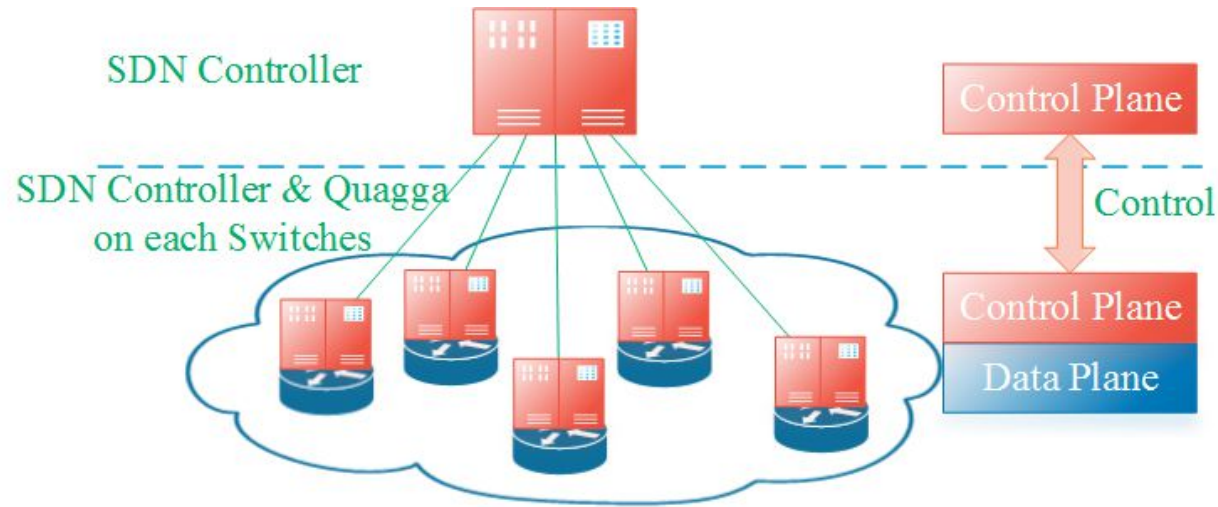
目的

- 段階的導入におけるAP-GW実装モデルの提案
 - 既存IPルーティングプロトコルへの対応
 - SDNコントローラの分散化によりスケーラビリティの向上



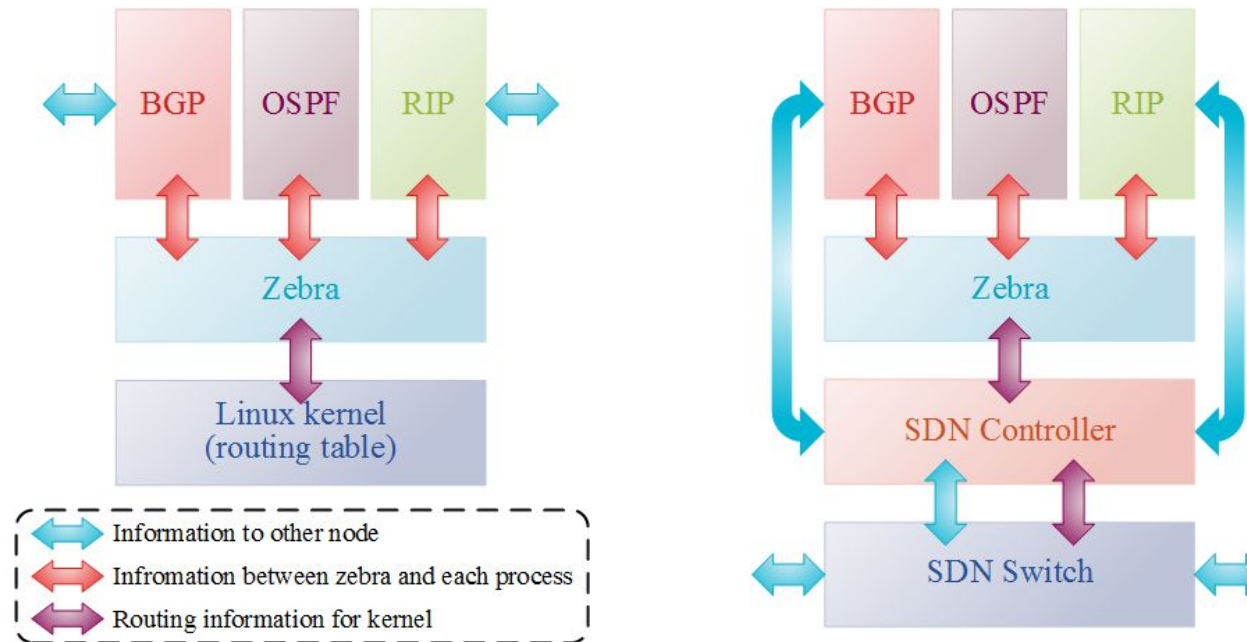
- 既存IP網との連携と、NWリソース管理機能の利用
 - 既存IPルーティングプロトコルの実装であるQuaggaを各SDNスイッチ上で動作させることで対応
 - SDNコントローラがSDNスイッチ上で動作することで処理を分散化
- 提案モデルのプロトタイプを実装
 - 提案モデルの実現可能性を評価

提案するAP-GWの実装モデル



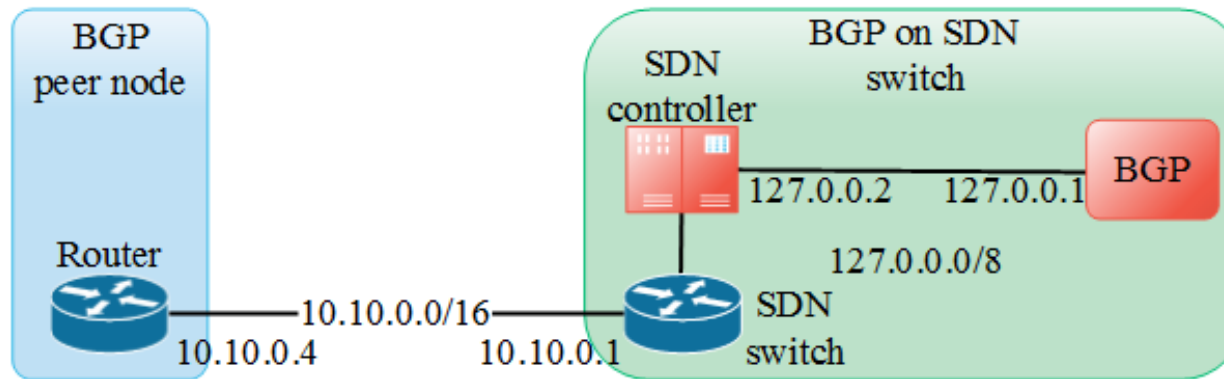
- スケーラビリティの向上
 - コントローラが分散化され, 処理が分散化可能
 - 大規模のネットワークにも対応可能
 - 既存ルーティングプロトコルへの対応による処理負荷の増加も軽減可
- 段階的導入も可能
 - フローテーブルとインターネットルーティングテーブルの同期管理が可能
 - スイッチ単体でのSDNの実現が可能となる
 - 集中管理だけがSDNの良さではない

各モジュールのSDNスイッチへの組み込み



- OFポートはOFの機能以外でパケットを送出できない
 - BGP などのデーモンが外部と通信するにはOFの機能が必要
 - ローカルの通信をパケットアウトする機能 (SDN-NAPT)が必要
- パケットの転送はフロー情報にもとづく
 - ルーティング情報とフロー情報を変換する機能が必要

SDN-NAPTの実装

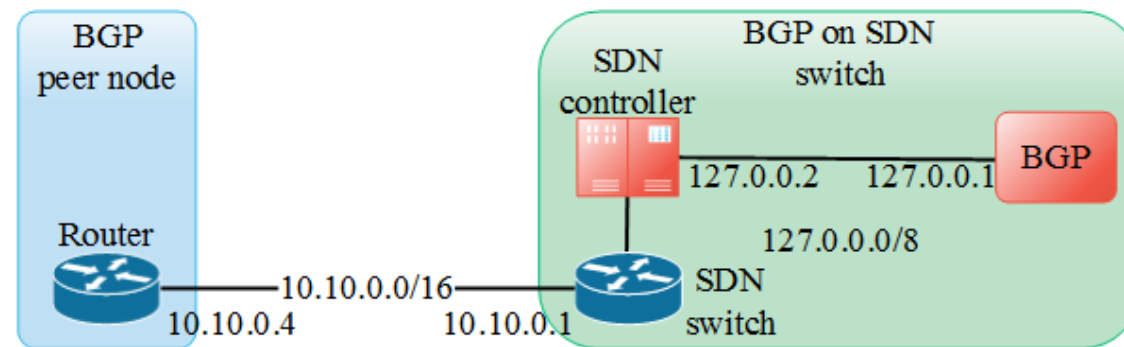


Global		Local (Loopback)	
dst	src	src	dst
10.10.0.4:179	10.10.0.1:37348	127.0.0.2:8179	127.0.0.1:44156

- 内部の通信をRAWソケットで，外部の通信に OFの機能を用いることでOFポートを用いた通信が可能
 - 通信の等価性を維持するためにRAWソケットを使用
 - BGPに限らず任意のサービスを稼働可能

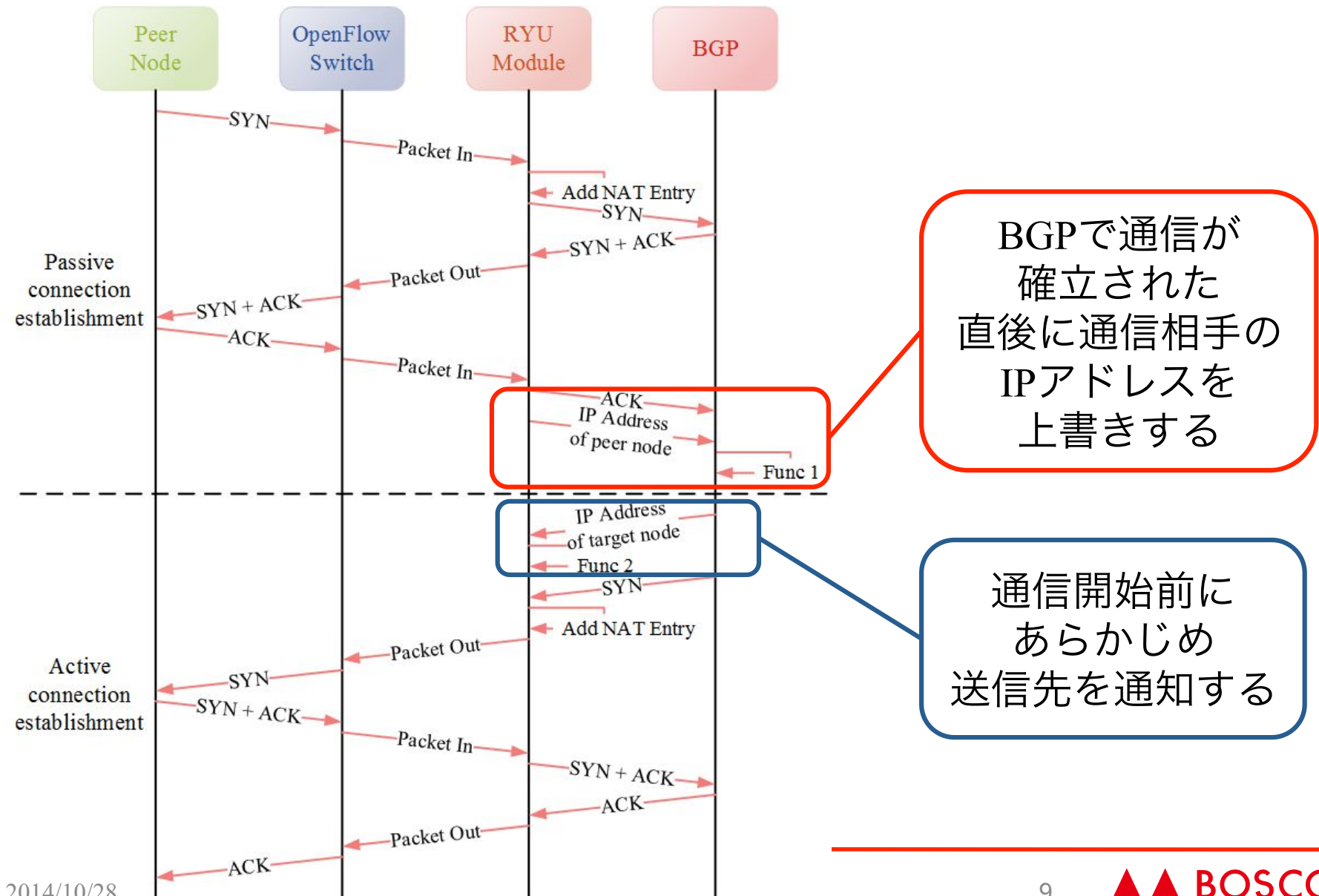
SDN-NAPT実装上の注意点

- ローカルのサービスから見ると、ゲートウェイはローカルアドレスとなる
 - ゲートウェイとして自分自身を指定できない



- BGPはピアノードとしか通信しない
 - BGPから見るとコントローラと直接通信することになるため設定にない接続と判断され、強制的に通信が切断される
- BGPはピア通信である
 - ローカル側から能動的にアクセスする場合、あらかじめコントローラに通信相手を通知する必要がある

SDN-NAPTにおけるセッション開始シーケンス



ルーティング情報とフロー情報の等価変換

- ルーティング情報とフローエントリの対比について

Kernel routing table	Flow entry
Destination	match : nw_dst
Gateway	actions : setfield
Metric	priority
Iface	actions : output

- 送信先はIPアドレスではなく、転送先のMACアドレスを使用
- priority と Metric は数値の意味が異なる

- 変換例

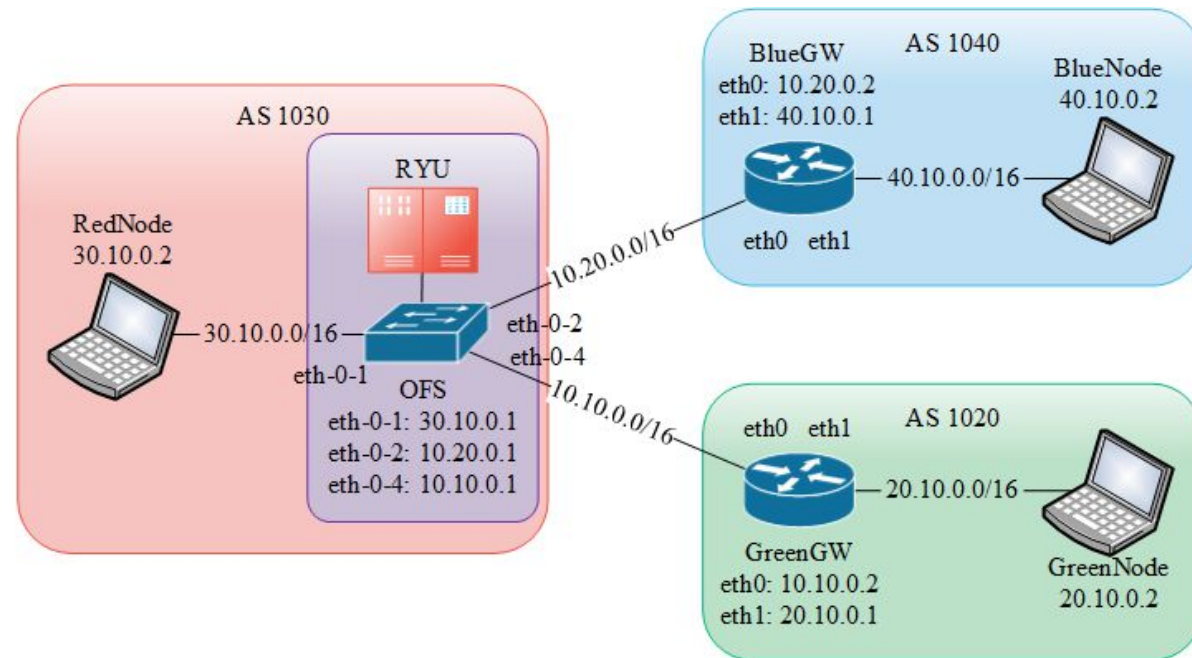
- ルーティング情報

```
# route -n
Kernel IP routing table
Destination    Gateway        Genmask         Flags Metric Ref    Use Iface
20.10.0.0      10.10.0.2      255.255.0.0     UG      0      0      0 eth1
```

- フローエントリ

```
# ovs-ofctl dump-flows br0 --protocol=OpenFlow13
OFPST_FLOW reply (OF1.3) (xid=0x2):
 cookie=0x0, duration=2.976s, table=0, n_packets=0, n_bytes=0,priority=1000,
 ip,nw_dst=20.10.0.0/16 actions=set_field:08:00:27:e0:db:bf->eth_dst,
 set_field:08:00:27:bf:22:e9->eth_src,output:2
```

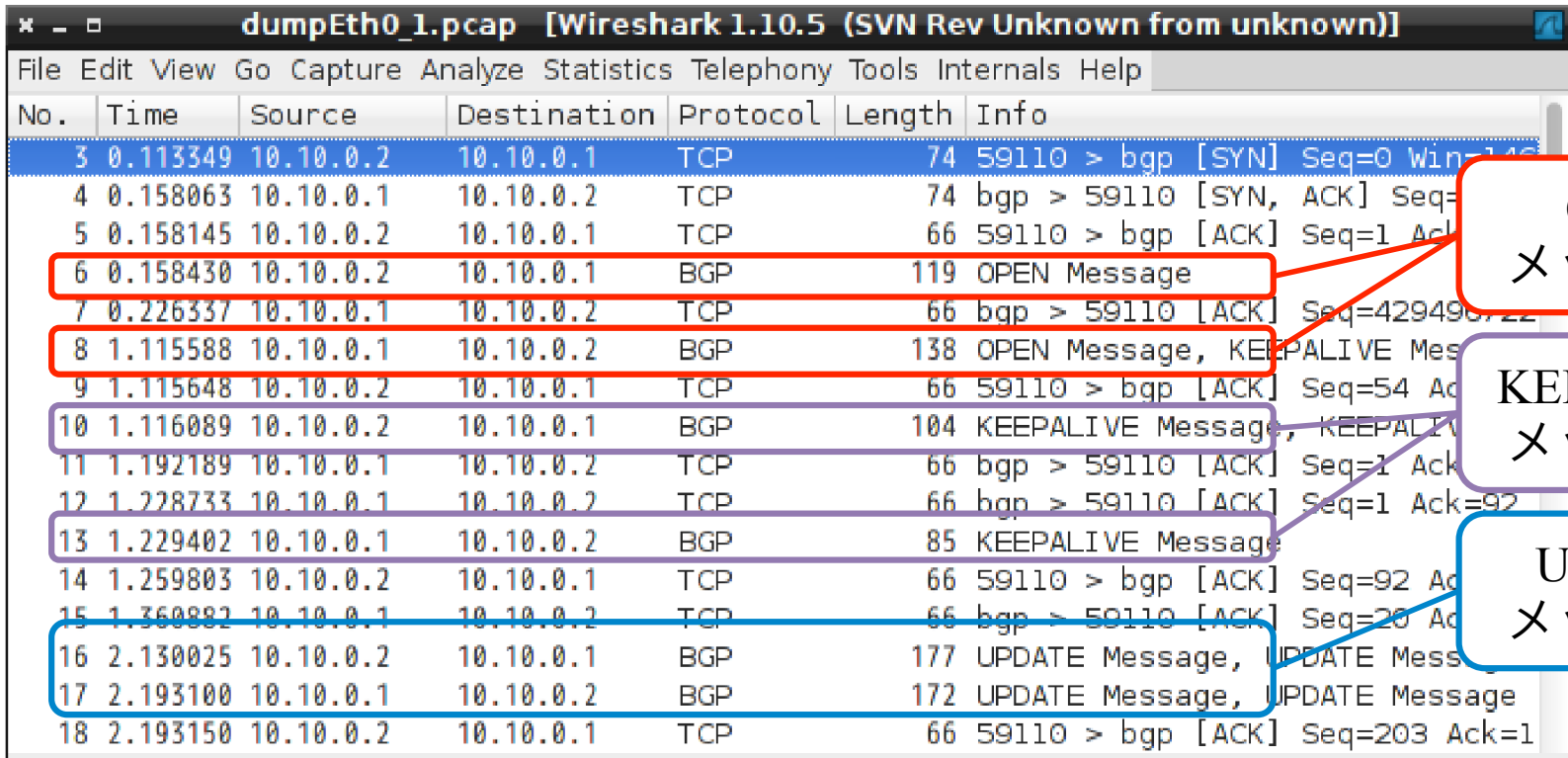
動作検証



- SDNスイッチの動作検証を行う
 - BGPメッセージの正常なやりとりの確認
 - BGPメッセージと連動した各ゲートウェイスイッチのフローおよびルーティング情報の確認
 - ping による疎通テスト

検証結果 (1)

- BGP メッセージの内容について
 - GreenGW, OFS 間の通信



No.	Time	Source	Destination	Protocol	Length	Info
3	0.113349	10.10.0.2	10.10.0.1	TCP	74	59110 > bgp [SYN] Seq=0 Win=1460
4	0.158063	10.10.0.1	10.10.0.2	TCP	74	bgp > 59110 [SYN, ACK] Seq=1460
5	0.158145	10.10.0.2	10.10.0.1	TCP	66	59110 > bgp [ACK] Seq=1460
6	0.158430	10.10.0.2	10.10.0.1	BGP	119	OPEN Message
7	0.226337	10.10.0.1	10.10.0.2	TCP	66	bgp > 59110 [ACK] Seq=429490722
8	1.115588	10.10.0.1	10.10.0.2	BGP	138	OPEN Message, KEEPALIVE Message
9	1.115648	10.10.0.2	10.10.0.1	TCP	66	59110 > bgp [ACK] Seq=54
10	1.116089	10.10.0.2	10.10.0.1	BGP	104	KEEPALIVE Message, KEEPALIVE Message
11	1.192189	10.10.0.1	10.10.0.2	TCP	66	bgp > 59110 [ACK] Seq=1
12	1.228733	10.10.0.1	10.10.0.2	TCP	66	bgp > 59110 [ACK] Seq=1 Ack=92
13	1.229402	10.10.0.1	10.10.0.2	BGP	85	KEEPALIVE Message
14	1.259803	10.10.0.2	10.10.0.1	TCP	66	59110 > bgp [ACK] Seq=92
15	1.360882	10.10.0.1	10.10.0.2	TCP	66	bgp > 59110 [ACK] Seq=20
16	2.130025	10.10.0.2	10.10.0.1	BGP	177	UPDATE Message, UPDATE Message
17	2.193100	10.10.0.1	10.10.0.2	BGP	172	UPDATE Message, UPDATE Message
18	2.193150	10.10.0.2	10.10.0.1	TCP	66	59110 > bgp [ACK] Seq=203 Ack=1

- 正しくBGPメッセージがやりとりされている

検証結果 (2)

- フローの登録内容
 - GreenGW

```
# route -n
Kernel IP routing table
Destination      Gateway         Genmask         Flags Metric Ref    Use Iface
0.0.0.0          10.10.0.1      0.0.0.0         UG    0      0      0 eth0
10.10.0.0        0.0.0.0        255.255.0.0     U     0      0      0 eth0
10.20.0.0        10.10.0.1      255.255.0.0     UG    1      0      0 eth0
20.10.0.0        0.0.0.0        255.255.0.0     U     0      0      0 eth1
30.10.0.0        10.10.0.1      255.255.0.0     UG    0      0      0 eth0
40.10.0.0        10.10.0.1      255.255.0.0     UG    0      0      0 eth0
```

AS1030
(OFS) の
経路情報

- OFS

```
Switch# ovs-ofctl dump-flows br0 --protocol=OpenFlow13
OFPST_FLOW reply (OF1.3) (xid=0x2):
  cookie=0x0, duration=216.25s, table=0, n_packets=0, n_bytes=0,
  priority=1000,ip,nw_dst=40.10.0.0/16 actions=set_field:00:0a:85:07:0c:34-
  >eth_dst,set_field:00:1e:08:08:93:38->eth_src,output:4
  cookie=0x0, duration=200.05s, table=0, n_packets=0, n_bytes=0,
  priority=1000,ip,nw_dst=20.10.0.0/16 actions=set_field:00:0a:85:07:0c:38-
  >eth_dst,set_field:00:1e:08:08:93:36->eth_src,output:2
  cookie=0x0, duration=218.471s, table=0, n_packets=31, n_bytes=2692,
  priority=10 actions=CONTROLLER:65535
```

AS1020
(Green) の
経路情報

経路情報が正しく交換されている
ping も経路情報にもとづき到達する

まとめと今後の課題

- まとめ

- 段階的に導入可能かつスケーラビリティが向上できるAP-GWの実装モデルを提案した
- 提案した実装モデルの実現のために必要なSDNスイッチの実装を行い、既存IPルーティングプロトコルへ対応可能であることを示した

- 今後の課題

- 大規模ネットワーク向け分散化コントローラの実装
- 実ネットワークの計測データを用いた評価および改良
- 最大フローエントリ数が少ないことに関する改善
 - スワップイン、スワップアウトによる対応の検討

