どこまで動く?
RPKI/Router

2012/7/6

Internet Multifeed Co. / JPNAP
Tomoya Yoshida

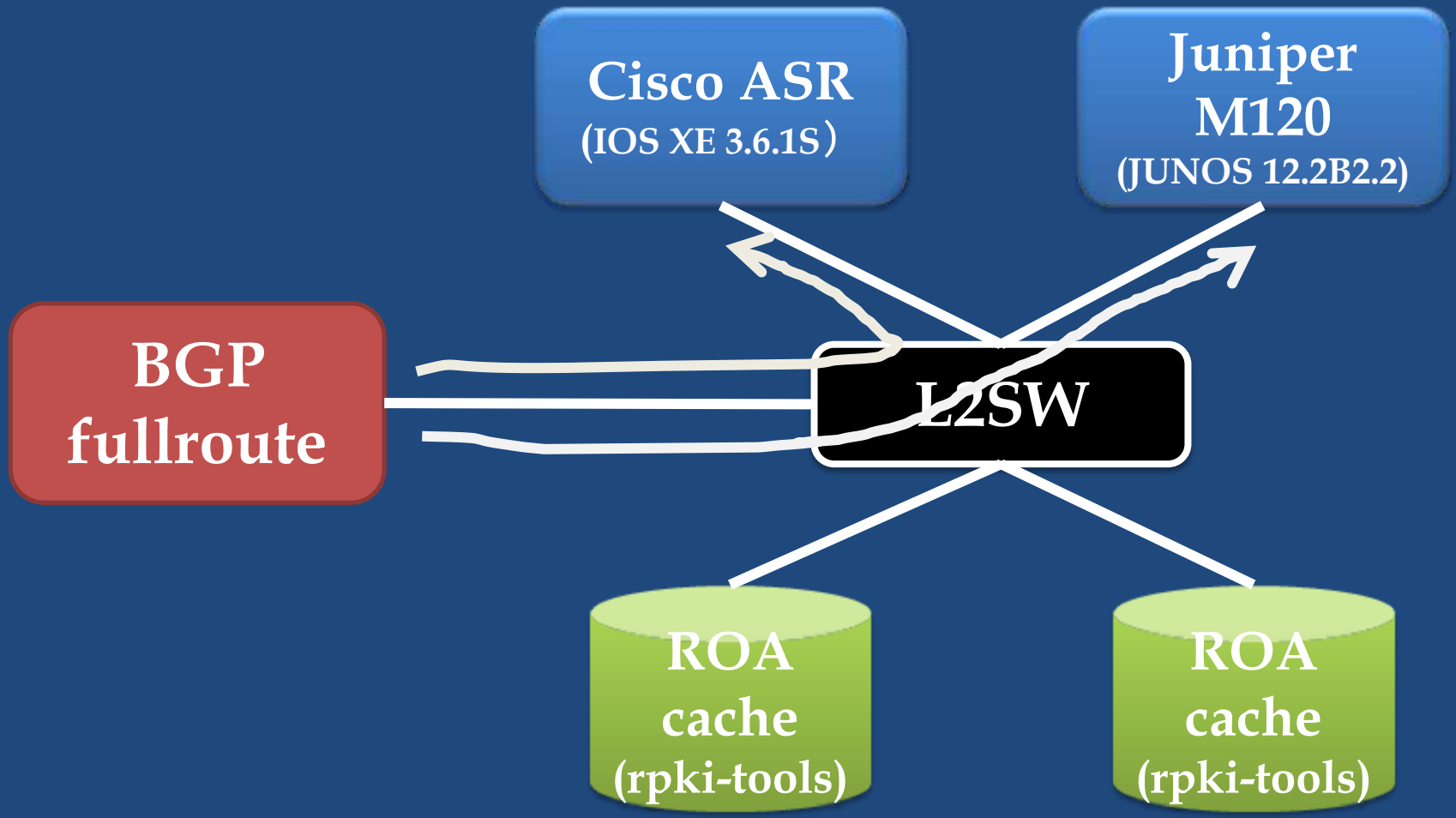
今日のトピック

- 実験結果の共有
- RPKI/Router周りの基本的な動き
- 今後の課題と展望

Cisco, Juniperで軽く実験してみました

- validation結果が想定通りになっているか
- Cisco, Juniperの差分
- 判定された経路のiBGP伝搬
- 45万ROA(v4)/1万ROA(v6)と愛し合えるか

実験環境



結果

- 基本的な動作はOK
- コマンドや表示結果の改良／追加が必要
- IBGPでCとJが愛し合えてない
  - 互いの努力は継続的に必要
- RPKIを動かさないほうが良いOSも…

結果のサマリ

	validation 結果	安定性	コマンドの豊富さ	ログ	パフォーマンス
Cisco ASR	○	？	△	○	？
Juniper M120 (beta)	○	？	◎	△	？

製品比較ではありません！！あくまでRPKI/Router protocolの今後の実装で、この辺が良くなると嬉しいなぁという観点で書いています！！

ルータがやること

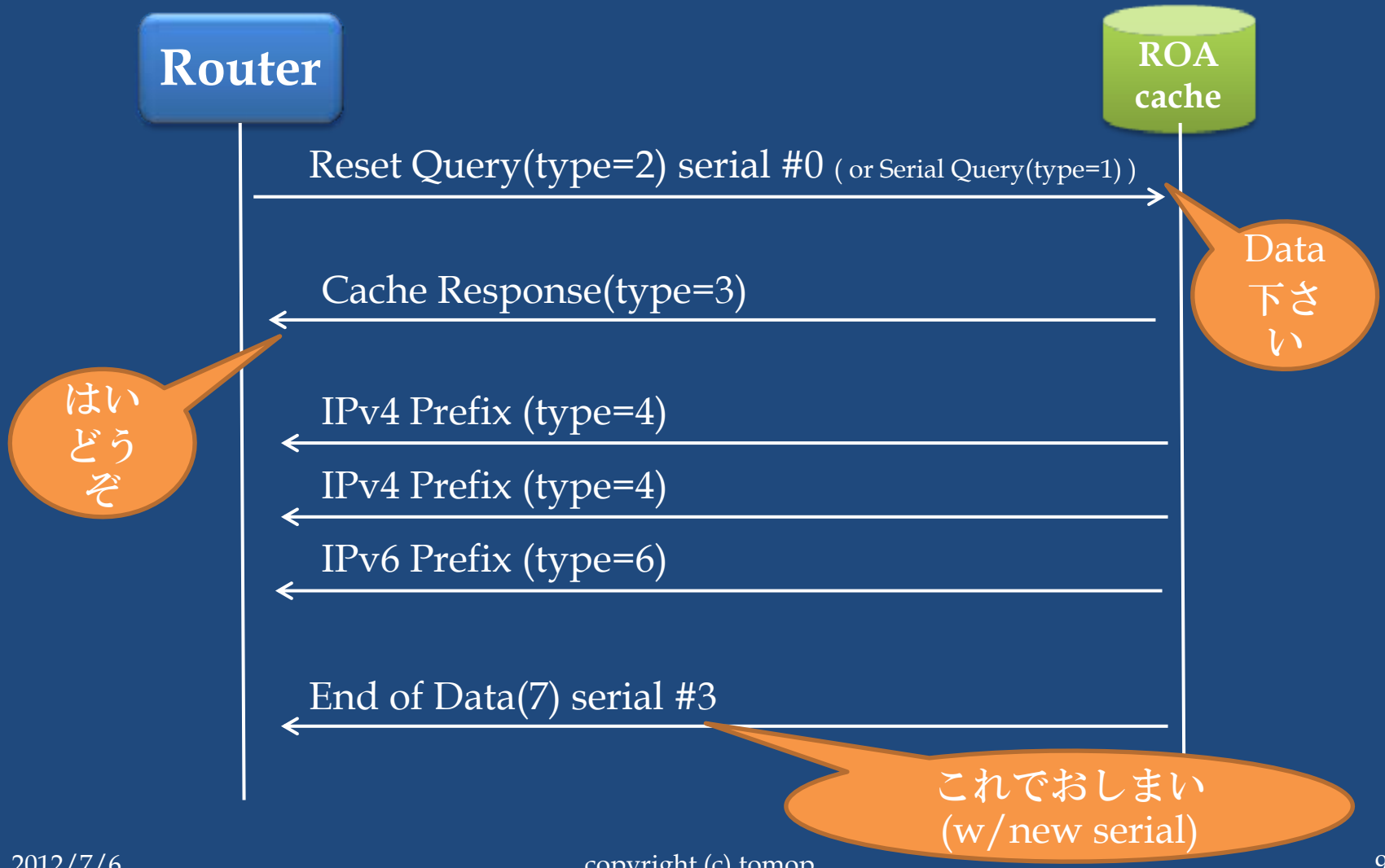
1. ルータにROAキャッシュ情報を蓄える
   - RTR protocolを用いて実施
2. 蓄えたROA情報を参照して、ルータ内でvalidation機構を動かす
3. marking されたRPKI validation status をiBGPで伝搬

ルータがやること

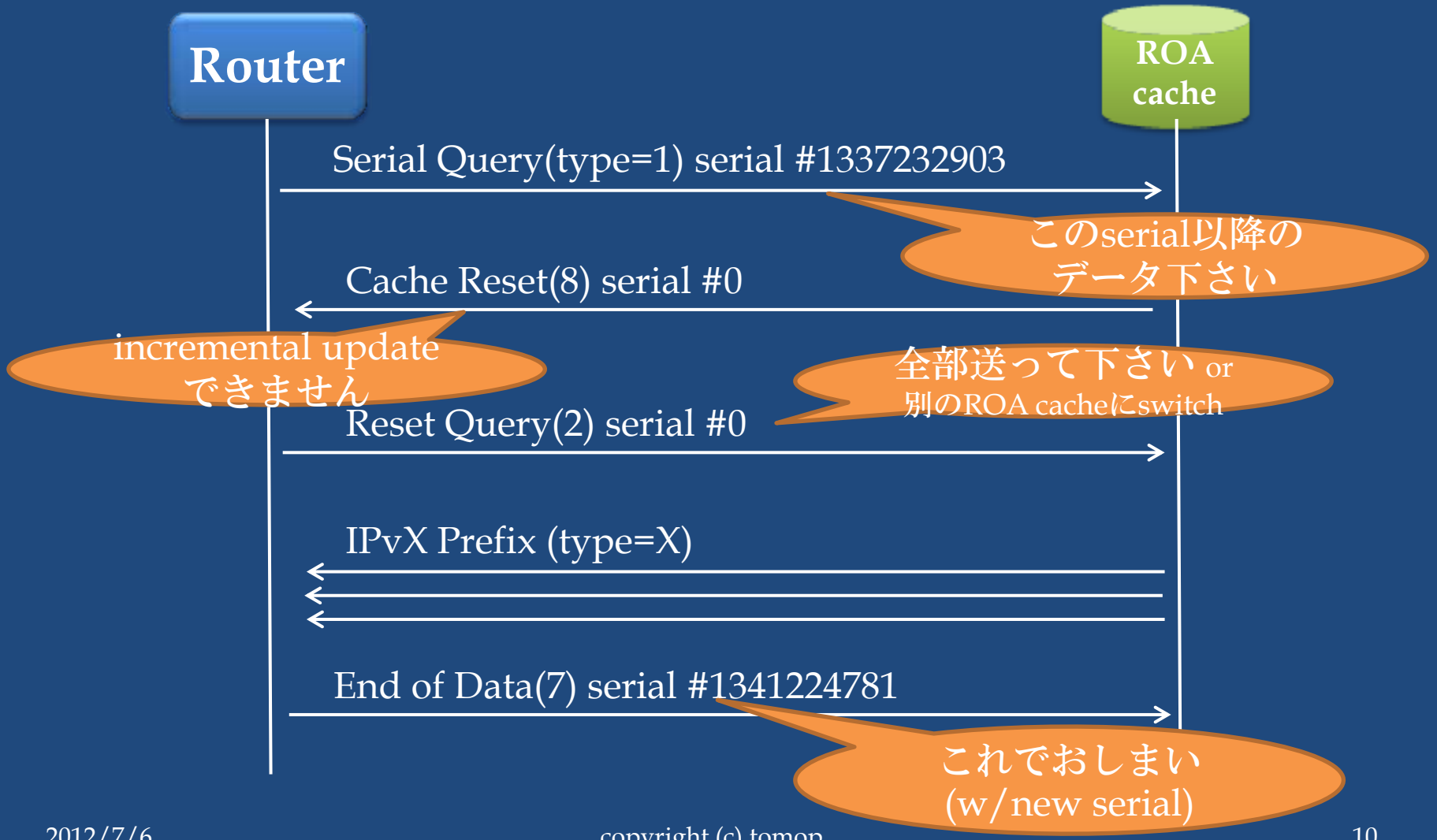
1. ルータにROAキャッシュ情報を蓄える
   - RTR protocolを用いて実施
2. 蓄えたROA情報を参照して、ルータ内でvalidation機構を動かす
3. marking されたRPKI validation status をiBGPで伝搬

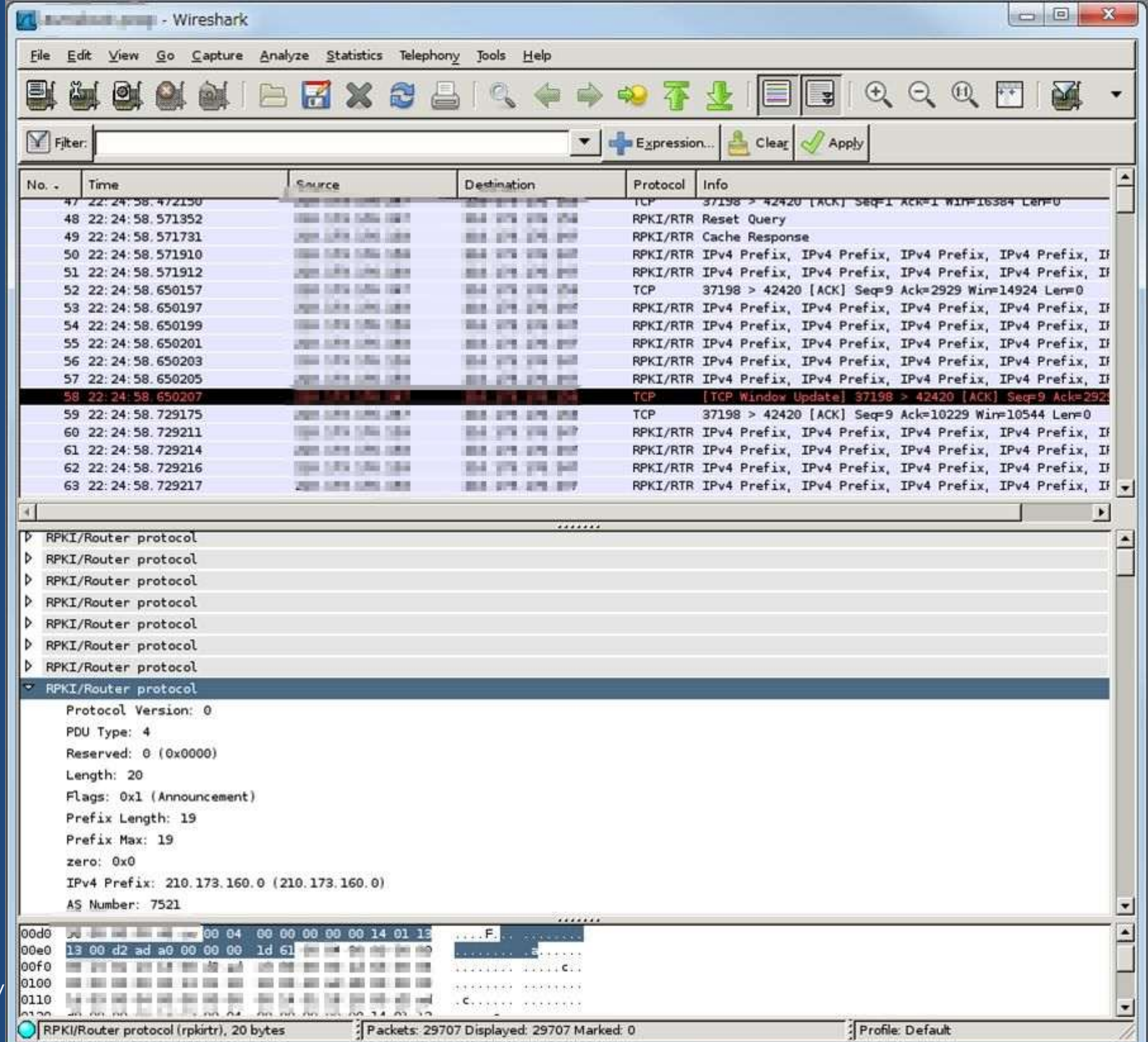
1. RTR(RPKI/Router) Protocol

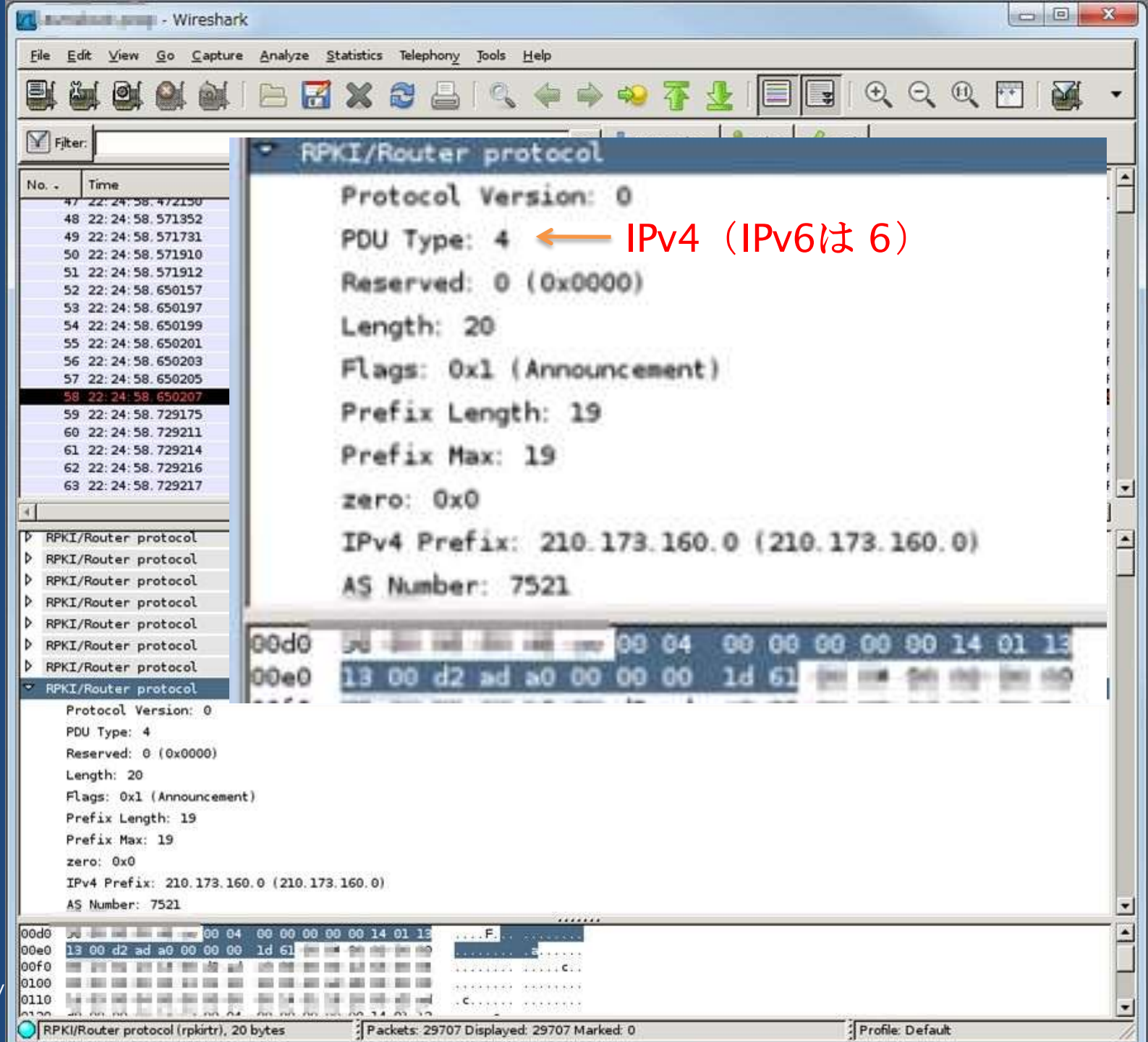
Start or Restart



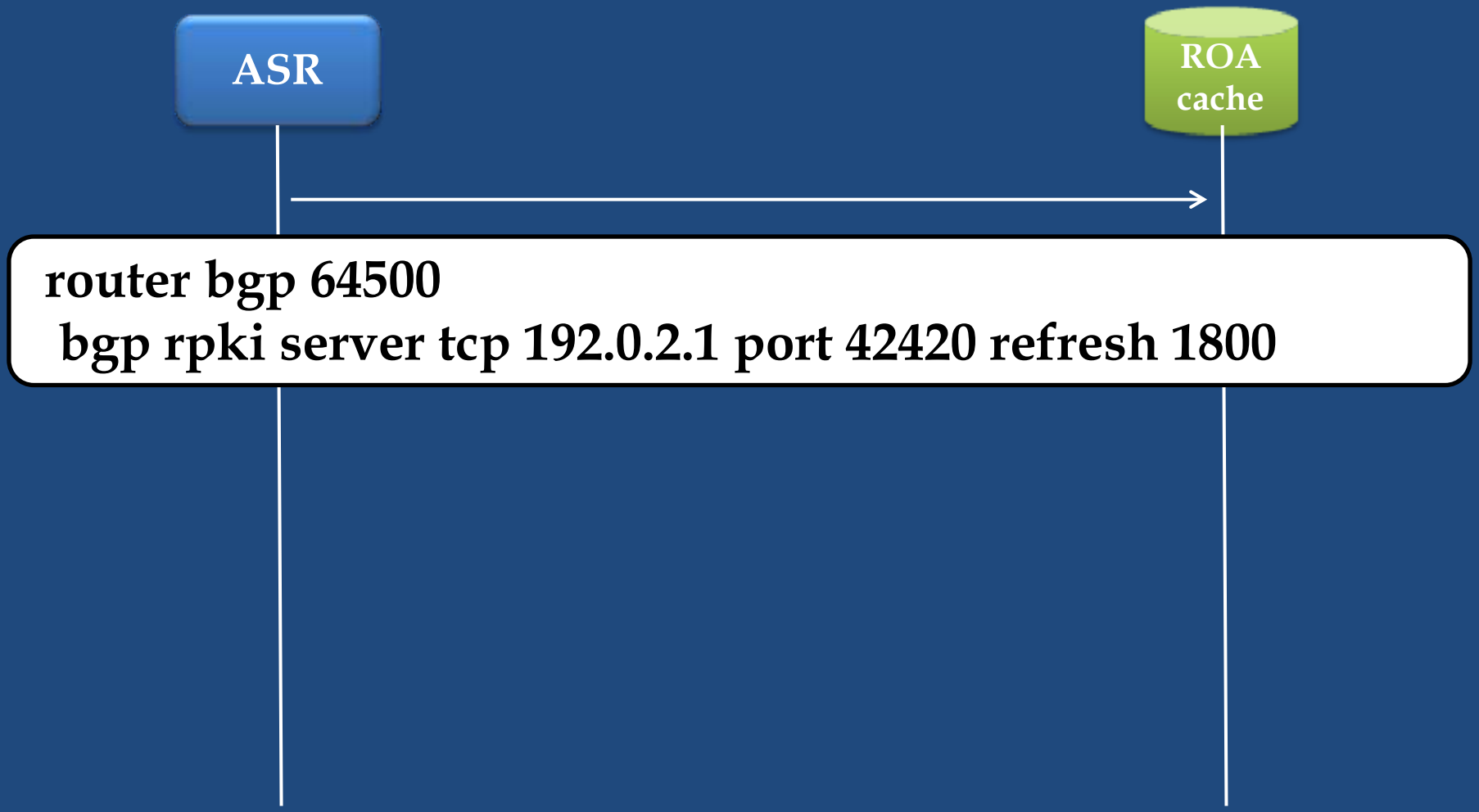
1. RTR(RPKI/Router) Protocol
update (no incremental update)




- Wireshark
File Edit View Go Capture Analyze Statistics Telephony Tools Help
Filter:
Expression... Clear Apply
No. Time Source Destination Protocol Info
48 22:24:58.571352 RPKI/RTR Reset Query
49 22:24:58.571731 RPKI/RTR Cache Response
50 22:24:58.571910 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
51 22:24:58.571912 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
52 22:24:58.650157 TCP 37198 > 42420 [ACK] Seq=9 Ack=2929 Win=14924 Len=0
53 22:24:58.650197 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
54 22:24:58.650199 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
55 22:24:58.650201 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
56 22:24:58.650203 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
57 22:24:58.650205 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
58 22:24:58.650207 TCP [TCP Window Update] 37198 > 42420 [ACK] Seq=9 Ack=292
59 22:24:58.729175 TCP 37198 > 42420 [ACK] Seq=9 Ack=10229 Win=10544 Len=0
60 22:24:58.729211 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
61 22:24:58.729214 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
62 22:24:58.729216 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
63 22:24:58.729217 RPKI/RTR IPv4 Prefix, IPv4 Prefix, IPv4 Prefix, IPv4 Prefix,
RPKI/Router protocol
RPKI/Router protocol
RPKI/Router protocol
RPKI/Router protocol
RPKI/Router protocol
RPKI/Router protocol
RPKI/Router protocol
RPKI/Router protocol
Protocol Version: 0
PDU Type: 4
Reserved: 0 (0x0000)
Length: 20
Flags: 0x1 (Announcement)
Prefix Length: 19
Prefix Max: 19
zero: 0x0
IPv4 Prefix: 210.173.160.0 (210.173.160.0)
AS Number: 7521
00d0 00 04 00 00 00 00 00 14 01 13
00e0 13 00 d2 ad a0 00 00 00 1d 61
00f0
0100
0110
RPKI/Router protocol (rpkirtr), 20 bytes
Packets: 29707 Displayed: 29707 Marked: 0
Profile: Default


Wireshark
File Edit View Go Capture Analyze Statistics Telephony Tools Help
Filter:
No. Time
48 22:24:58.571352
49 22:24:58.571731
50 22:24:58.571910
51 22:24:58.571912
52 22:24:58.650157
53 22:24:58.650197
54 22:24:58.650199
55 22:24:58.650201
56 22:24:58.650203
57 22:24:58.650205
58 22:24:58.650207
59 22:24:58.729175
60 22:24:58.729211
61 22:24:58.729214
62 22:24:58.729216
63 22:24:58.729217
RPKI/Router protocol
Protocol Version: 0
PDU Type: 4
Reserved: 0 (0x0000)
Length: 20
Flags: 0x1 (Announcement)
Prefix Length: 19
Prefix Max: 19
zero: 0x0
IPv4 Prefix: 210.173.160.0 (210.173.160.0)
AS Number: 7521
IPv4（IPv6は 6）
00d0 00 04 00 00 00 00 00 14 01 13
00e0 13 00 d2 ad a0 00 00 00 1d 61
RPKI/Router protocol (rpkirtr), 20 bytes
Packets: 29707 Displayed: 29707 Marked: 0
Profile: Default

1. RTR(RPKI/Router) Protocol



1. RTR(RPKI/Router) Protocol


```
routing-options {
 validation {
 group ROA {
        session 192.0.2.1 {
            refresh-time 1800;
            port 42420;
            local-address 192.0.2.13;
        }
    }
}
```

ASR

asr>show bgp ipv4 unicast rpki table

Load for five secs: 0%/0%; one minute: 1%; five minutes: 1%
Time source is NTP, 11:26:41.011 JST Fri Jul 6 2012
452768 BGP sovc network entries using 72442880 bytes of memory
455551 BGP sovc record entries using 14577632 bytes of memory

Network	Maxlen	Origin-AS	Source	Neighbor
1.0.0.0/24	24	15169	0	192.0.2.1/42420
1.0.4.0/22	22	56203	0	192.0.2.1/42420
1.0.16.0/23	23	2519	0	192.0.2.1/42420
1.0.18.0/23	23	2519	0	192.0.2.1/42420
1.0.20.0/23	23	2519	0	192.0.2.1/42420
1.0.22.0/23	23	2519	0	192.0.2.1/42420
1.0.24.0/24	24	2519	0	192.0.2.1/42420
1.0.24.0/23	23	2519	0	192.0.2.1/42420
1.0.25.0/24	24	2519	0	192.0.2.1/42420
1.0.26.0/24	24	2519	0	192.0.2.1/42420

ASR

asr>show bgp ipv6 unicast rpki table

Load for five secs: 2%/0%; one minute: 6%; five minutes: 2%
Time source is NTP, 10:56:34.272 JST Fri Jul 6 2012
9851 BGP sovc network entries using 1812584 bytes of memory
9932 BGP sovc record entries using 317824 bytes of memory

Network	Maxlen	Origin-AS	Source	Neighbor
2001::/32	32	1101	0	192.0.2.1/42420
2001::/32	32	6939	0	192.0.2.1/42420
2001::/32	32	12859	0	192.0.2.1/42420
2001:200::/32	32	2500	0	192.0.2.1/42420
2001:200:900::/40	40	7660	0	192.0.2.1/42420
2001:200:905::/48	48	56218	0	192.0.2.1/42420
2001:200:C00::/40	40	7530	0	192.0.2.1/42420
2001:200:C000::/35	35	23634	0	192.0.2.1/42420
2001:200:E000::/35	35	7660	0	192.0.2.1/42420

ASR

asr>show ip bgp rpki ?
  servers Display RPKI cache server information
  table   Display RPKI table entries

statistics関連のshowコマンドの充実、もしくは表示上見やすくして貰えるとうれしいです。

M120

m120> show validation database session 192.0.2.2

RV database for instance master

Prefix	Origin-AS	Session	State	Mismatch
210.173.160.0/19-24	7521	192.0.2.2	valid	
2001:3a0::/32-64	7521	192.0.2.2	valid	

IPv4 records: 1

IPv6 records: 1

M120

m120> show validation database session 192.0.2.1

RV database for instance master

Prefix	Origin-AS	Session	State	Mismatch
1.0.0.0/24-24	15169	192.0.2.1	valid	
1.0.4.0/22-22	56203	192.0.2.1	valid	
1.0.16.0/23-23	2519	192.0.2.1	valid	
1.0.18.0/23-23	2519	192.0.2.1	valid	
1.0.20.0/23-23	2519	192.0.2.1	valid	
1.0.22.0/23-23	2519	192.0.2.1	valid	
1.0.24.0/23-23	2519	192.0.2.1	valid	
1.0.24.0/24-24	2519	192.0.2.1	valid	
1.0.25.0/24-24	2519	192.0.2.1	valid	
1.0.26.0/23-23	2519	192.0.2.1	valid	

IPv4フルルートが終わった後にIPv6 ROAが表示される（検索が苦しい…）

M120

m120> show validation database origin-autonomous-system 7521
RV database for instance master

Prefix	Origin-AS	Session	State	Mismatch
210.173.160.0/19-19	7521	192.0.2.1	valid	
210.173.160.0/19-24	7521	192.0.2.2	valid	
2001:3a0::/32-32	7521	192.0.2.1	valid	
2001:3a0::/32-64	7521	192.0.2.2	valid	

IPv4 records: 2
IPv6 records: 2

OriginAS単位で検索が可能（good）

M120

m120> show validation session

Session	State	Flaps	Uptime	#IPv4/IPv6 records
192.0.2.1	Up	0	00:45:05	455550/9931
192.0.2.2	Up	0	17:25:52	1/1

表示がBGPのsummry経路数表示に似ていてなじみやすい

1. RTR(RPKI/Router) Protocol

- CiscoとJuniperの違い
  - Cisco ASR
    - 設定後、ROA cacheがルータ内に生成
    - Validationが行われる
      - Valid, Not found：BGP tableへ
      - Invalid：defaultの状態では BGP tableへ載らない
  - Juniper M120
    - 設定後、ROA cacheがルータ内に生成
    - 全てunverified状態となる

ルータがやること

1. ルータにROAキャッシュ情報を蓄える
   - RTR protocolを用いて実施
2. 蓄えたROA情報を参照して、ルータ内でvalidation機構を動かす
3. marking されたRPKI validation status をiBGPで伝搬

Validation status

1. Valid	origin AS, prefix, 最大prefix長がROAの範囲に存在する場合
2. Not found (unknown)	合致するprefixとprefix長をもつROAが存在しない場合
3. Invalid	prefix, 許可された最大prefix長が合致するROAは存在するが、originASが異なる場合

(1)OriginAS (2)Prefix (3)max prefix length の3つに基づいてvalidationする

2. Validation (ASR)

```
router bgp 64500
 address-family ipv4
  bgp bestpath prefix-validate allow-invalid

route-map rpki permit 10
 match rpki invalid
 set community 65400:2 additive
!
route-map rpki permit 20
 match rpki not-found
 set community 65400:1 additive
!
route-map rpki permit 30
 match rpki valid
 set community 65400:0 additive
```

Invalid経路もbestpath selectionに反映

左記のようなpolicy設定が不要なら本設定は不要

Validationの結果に応じて以下のExtended Communityが自動付与される

Valid	0x43:0:0
unkown	0x43:0:1
Invalid	0x43:0:2

2. Validation (M120)

```
protocols {
    bgp {
      group RPKI-fullroute {
          neighbor 192.0.2.254 {
              import validation;
              peer-as 131079;
          }
          neighbor 2001:7fa:7:1:0:13:1079:1 {
              import validation;
              peer-as 131079;
          }
      }
```

2. Validation (M120)

```
policy-statement validation {
    term valid {
        from {
            protocol bgp;
            validation-database valid;
        }
        then {
            validation-state valid;
            community set rpki-valid;
            community add origin-validation-state-valid;
            accept;
        }
    }
    term invalid {
        from {
            protocol bgp;
            validation-database invalid;
        }
        then {
            validation-state invalid;
            community set rpki-invalid;
            community add origin-validation-state-invalid;
            accept;
        }
    }
    term unknown {
        from {
            protocol bgp;
            validation-database unknown;
        }
        then {
            validation-state unknown;
            community set rpki-unknown;
            community add origin-validation-state-unknown;
            accept;
        }
    }
}

 community origin-validation-state-invalid members 0x43:65400:2;
 community origin-validation-state-unknown members 0x43:65400:1;
 community origin-validation-state-valid members 0x43:65400:0;
 community rpki-invalid members 65400:2;
 community rpki-unknown members 65400:1;
 community rpki-valid members 65400:3;
```

M120(IPv4)

```
asr>show ip bgp 210.173.160.0/19
Load for five secs: 0%/0%; one minute: 1%; five minutes: 1%
Time source is NTP, 10:54:10.058 JST Fri Jul 6 2012
BGP routing table entry for 210.173.160.0/19, version 938277
Paths: (2 available, best #1, table default)
  Not advertised to any peer
  Refresh Epoch 1
  131079 7521
    192.0.2.254 from 192.0.2.254 (210.173.172.118)
      Origin IGP, localpref 100, valid, external, best
      Community: 65400:1
      path 7FC93CD9B9C0 RPKI State valid
  Refresh Epoch 1
  131079 7521, (received-only)
    192.0.2.254 from 192.0.2.254 (210.173.172.118)
      Origin IGP, localpref 100, valid, external
      path 7FC93CD9B958 RPKI State valid
```

ASR(IPv6)

```
asr>show bgp ipv6 uni 2001:3a0::/32
Load for five secs: 0%/0%; one minute: 1%; five minutes: 1%
Time source is NTP, 10:52:19.542 JST Fri Jul 6 2012
BGP routing table entry for 2001:3A0::/32, version 16909
Paths: (2 available, best #2, table default)
  Advertised to update-groups:
     2
  Refresh Epoch 1
  131079 7521
    2001:7FA:7:1:0:13:1079:1 from 2001:7FA:7:1::250:1 (210.173.161.247)
      Origin IGP, localpref 100, valid, internal
      Community: 65400:3
      Extended Community: 0x43:65400:0
      path 7FC93120BC28 RPKI State valid
  Refresh Epoch 1
  131079 7521, (received & used)
    2001:7FA:7:1:0:13:1079:1 (FE80::205:85FF:FE16:C001) from 2001:7FA:7:1:0:13:1079:1
      (210.173.172.118)
      Origin IGP, localpref 100, valid, external, best
      path 7FC931062E08 RPKI State valid
```

Validation結果（1/7）

ROA

BGP

10.0.0.0/16-16	AS65000
10.0.0.0/16	AS65000

Valid

10.0.0.0/16-16	AS65000
10.0.0.0/16	AS65001

Invalid

Validation結果（2/7）

ROA

BGP

10.0.0.0/16	AS65000

Not Found

10.0.0.0/16	AS65001

Not Found

Validation結果（3/7）

ROA
BGP

10.0.0.0/16-16 AS65000

10.0.0.0/8 AS65000

Not Found

10.0.0.0/16-16 AS65000

10.0.0.0/17 AS65000

Invalid

10.0.0.0/16-24 AS65000

10.0.0.0/17 AS65000

Valid

Validation結果（4/7）

ROA
BGP

10.0.0.0/16-16	AS65000
10.0.0.0/16-16	AS65001
10.0.0.0/16	AS65000

Valid

10.0.0.0/16-16	AS65000
10.0.0.0/16-16	AS65001
10.0.0.0/16	AS65001

Valid

Validation結果（5/7）

ROA
BGP

10.0.0.0/17-17 AS65000

10.0.128.0/17-17 AS65000

10.0.0.0/16 AS65000

Not Found

Validation結果（6/7）

ROA
BGP

10.0.0.0/16-24 AS0
10.0.0.0/8 AS65000

Not Found

10.0.0.0/16-24 AS0
10.0.0.0/24 AS65000

Invalid

10.0.0.0/16-24 AS0
10.0.0.0/32 AS65000

Invalid

Validation結果（7/7）

ROA
BGP

10.0.0.0/16-24　AS65000
10.0.0.0/24　{AS65000}

Not Found

10.0.0.0/16-24　AS65000
10.0.0.0/24　{AS65001}

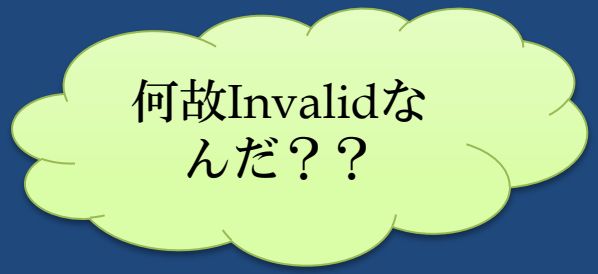
Not Found

10.0.0.0/16-24　AS65000
10.0.0.0/24　{AS65000, AS65001}

Not Found

とあるPrefixの例

```
asr#show ip bgp 109.5.117.0/24
BGP routing table entry for 109.5.117.0/24, version 231295
131079 7521 2497 15557 41334
    192.0.2.254 from 192.0.2.254 (210.173.172.118)
      Origin IGP, localpref 100, valid, external, best
      Community: 65400:1
      path 7F9B26F111D0 RPKI State invalid
  Refresh Epoch 1
  131079 7521 2497 15557 41334, (received-only)
    192.0.2.254 from 192.0.2.254 (210.173.172.118)
      Origin IGP, localpref 100, valid, external
      path 7F9B26F11168 RPKI State valid
```



とあるPrefixの例

- ROA全ルートを上からたどって見つけるしかない…

とあるPrefixの例

＃ROA

・・・

109.0.0.0/11　　11　　15557　　0　　210.173.176.117/42420

109.0.0.0/11-11　AS15557

109.5.117.0/24 AS41334

運用上の課題

- Validation結果が、何のROAに基づいているのかを検索するのが困難
  - BGPのPrefixを包含するROAを探すか、それが無い事を探す必要がある
- Reasonを経路情報に記述するか、コマンドを叩いて検索できるようにしてほしい

ルータがやること

1. ルータにROAキャッシュ情報を蓄える
   - RTR protocolを用いて実施
2. 蓄えたROA情報を参照して、ルータ内でvalidation機構を動かす
3. marking されたRPKI validation status をiBGPで伝搬

iBGPへの伝搬

- 基本的にはborder routerがRPKI origin validationを実施し、内部のルータへRPKI validation statusを伝搬
  - Extended Communityを経路受信時につけて、iBGPでstatusを他のルータに伝搬
- Extended Communityの実装がCiscoとJuniperで異なることが発覚

Extended Communityの実装

	Valid	Not found	Invalid
Cisco ASR	0x43:0:0	0x43:0:1	0x43:0:2
Juniper M120	0x43:X:0	0x43:X:1	0x43:X:2

Juniperの場合は、X: AS番号

今回の検証では愛し合えませんでした。。

45万(v4)+1万(v6)ROAに挑戦

- RISのBGP経路に基づきROAを作成
- ルータにimport

45万(v4)+1万(v6)ROAに挑戦

※M120の結果はJUNOS 12.2B2.2（ベータコードによる試験）

	ROAを受けきる時間
Cisco ASR	20秒
Juniper M120	4分25秒

IPv4:41万 IPv6:9300 経路を保有している状態

ROA cache server 次第で結果が異なる可能性も高い。またルータのパフォーマンス次第なのであくまで1結果として記述している（以降も同様）

45万(v4)+1万(v6)ROAに挑戦

	IPv4 fullrouteを受けきる時間	IPv6 fullrouteを受けきる時間
Cisco ASR	1分2秒	3秒
Juniper M120	2分35秒	9秒

IPv4:45万 IPv6:9900 ROAを保有している状態

45万(v4)+1万(v6)ROAに挑戦

	IPv4 fullrouteを受けきる時間	IPv6 fullrouteを受けきる時間
Cisco ASR	1分4秒	1秒
Juniper M120	2分34秒	12秒

ROAを保有していない通常状態

その他の課題

- RTRセッションが切れてしばらくすると、全てnot foundになる
  - RTRセッションを複数確立し、それらが切れないことが大前提
  - Invalid => not found
    - Invalid状態でpolicyにて無効化されていたものが、not found状態に遷移することで、他にUpdateされる可能性が十分にある => hijacking

雑感

- RTR、なんとかいけるかな。。
  - 基本的な機能は出来上がってきた
- まだまだ細かい挙動の確認が必要
- RPKI/Router Protocolについては、エンジニアはきちんと挙動を理解しておかないといけない

今後の課題と展望

- ルーティングシステム全体の影響をきちんと考える必要あり
  - ルータが不意にreloadした際の挙動
    - BGP経路受信とROAキャッシュのタイミングによっては、経路選択方法に差異が発生する
  - ROA cacheとの信頼関係
    - レジストリのROA運用
- 今後のdeployment
  - 局所的にやってもあまり効果がない
    - 上流ISPへ流れた瞬間Invalidの経路に向いてしまう可能性が十分にある
  - 守りたいもの vs リスク を真剣に考える必要あり