

SRv6@ShowNetとプログラマビリティ

東京大学/ShowNet NOCチームメンバー
中村 遼

Open Networking Conference 2022

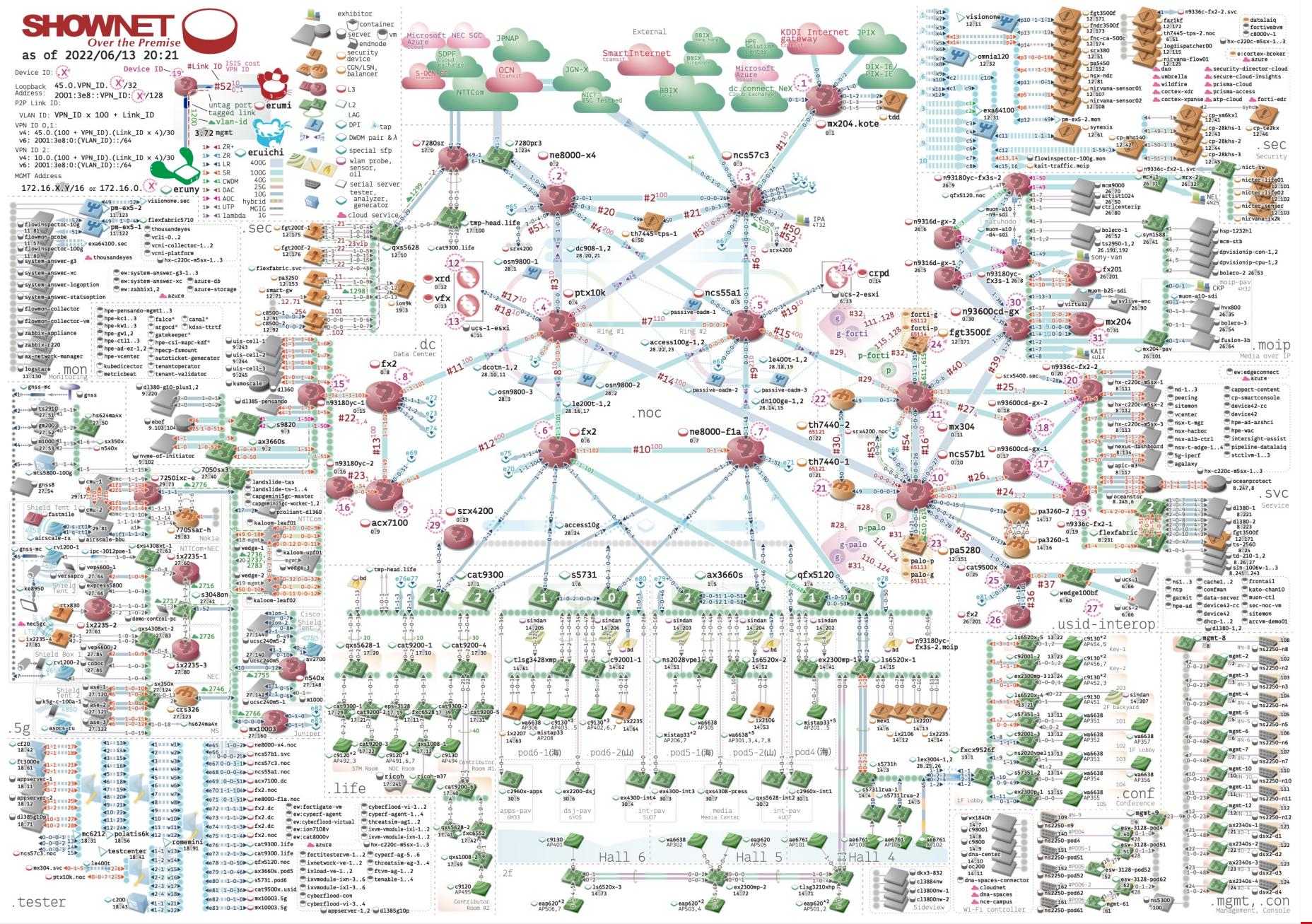
自己紹介

- 東京大学
 - 情報基盤センター ネットワーク研究部門 准教授
 - ネットワークアーキテクチャ、OSネットワークサブシステム
 - 東京大学キャンパスネットワークの運用
- Interop Tokyo ShowNet NOCチームメンバー
 - 主にL2/L3を担当
 - SDNとかもやっていました

Interop Tokyo **SHOWNET** *Over the Premise*

- **Interop:** 世界最大のネットワーク機器と技術の展示会
 - 毎年6月に幕張メッセで開催、2022年は来場者約9万人
- **ShowNet:** 世界最大のデモンストレーションネットワーク
 - その年々の最新技術/プロトコルの相互接続性検証やデモを実施
 - 出展社や来場者へのネットワーク接続性を提供するLive Network





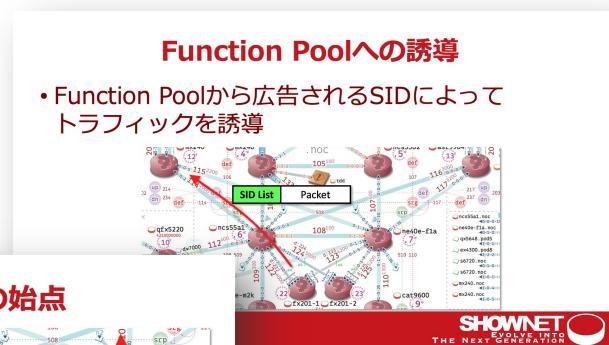
ShowNetのアイコンは
Interop Tokyoの
ShowNet 2022のページから
フリー (CC-BY-SA 4.0) で
ダウンロードできます！



<https://www.interop.jp/2022/shownet/concept/>
「ShowNetアイコン」の画像をクリック

ShowNetにおけるSegment Routingの取り組み

- 2019年: SRv6 Service Chaining
 - 6機種5ベンダーのSRv6機器で構築
 - draft-upa-srv6-service-chaining-exp, draft-eden-srv6-tagging-proxy



ShowNetにおけるSegment Routingの取り組み

- 2021年: SR-MPLS/SRv6バックボーン
 - SR-MPLS: Flex-Algo, BGP Egress Peer Engineering
 - SRv6: draft-ietf-bess-srv6-servicesベースの相互接続検証

APNIC

BLOG SEARCH ADVANCED WHOIS MAKE A PAYMENT Your IP address: LOG IN

Get IP Manage IP Training Events Insights Community Blog Help Centre About Contact

Measuring the potential benefit of egress traffic engineering with Segment Routing
By Ryo Nakamura on 10 Mar 2022 Category: Tech matters Tags: BGP, Guest Post, Japan, measurement, segment routing

Photo: Nic McPhee, Flickr.

Border Gateway Protocol (BGP) best paths on which the Internet relies are not always the 'best' when it comes to bandwidth and latency. This is due to several factors including BGP not considering congestion on links and AS-PATHs using prepending no longer reflect closeness.

As such, engineering egress traffic toward neighboring Autonomous Systems (ASes) is a constant challenge for network operators seeking better paths.

To begin with, the potential benefit of egress traffic engineering is unclear. Measuring the performance of a non-best path to a destination requires preferring the non-best path. However, using local preference to do this impacts all user traffic to the

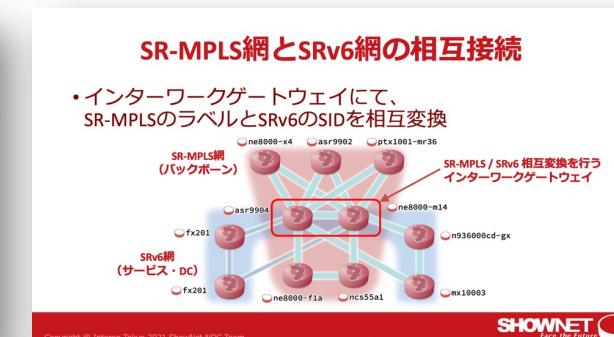
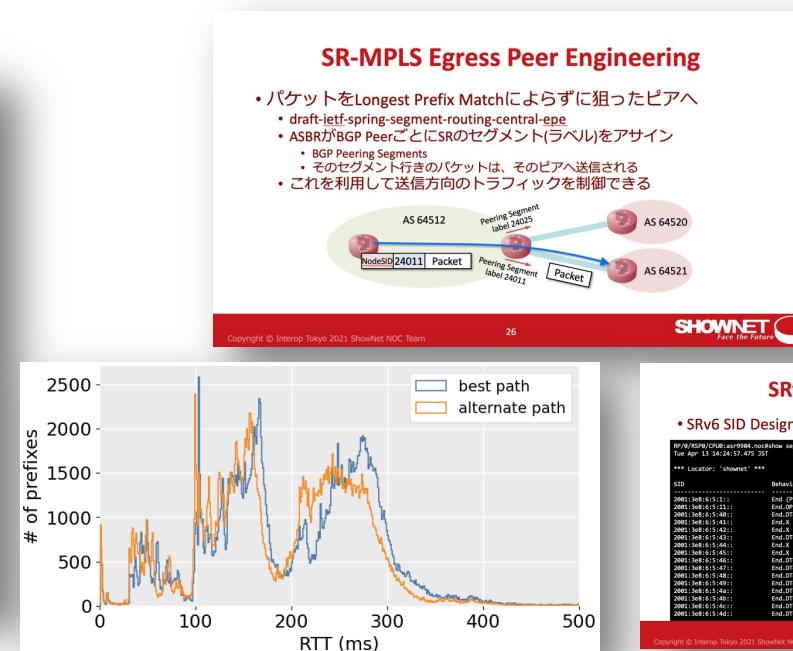
Get Updates Email * Subscribe! Show options Select list(s): Daily Weekly

Latest Tweets Learn how a new #OpenSource #RPC dial-out collector is helping one ISP add more data streams to its telemetry solution. <https://t.co/rwI073abSA>

Introducing Orbit, an evolution of mailing lists. Paul Wilson, @apnicdo, shares the

of prefixes 0 500 1000 1500 2000 2500 0 100 200 300 400 500 RTT (ms)

best path alternate path



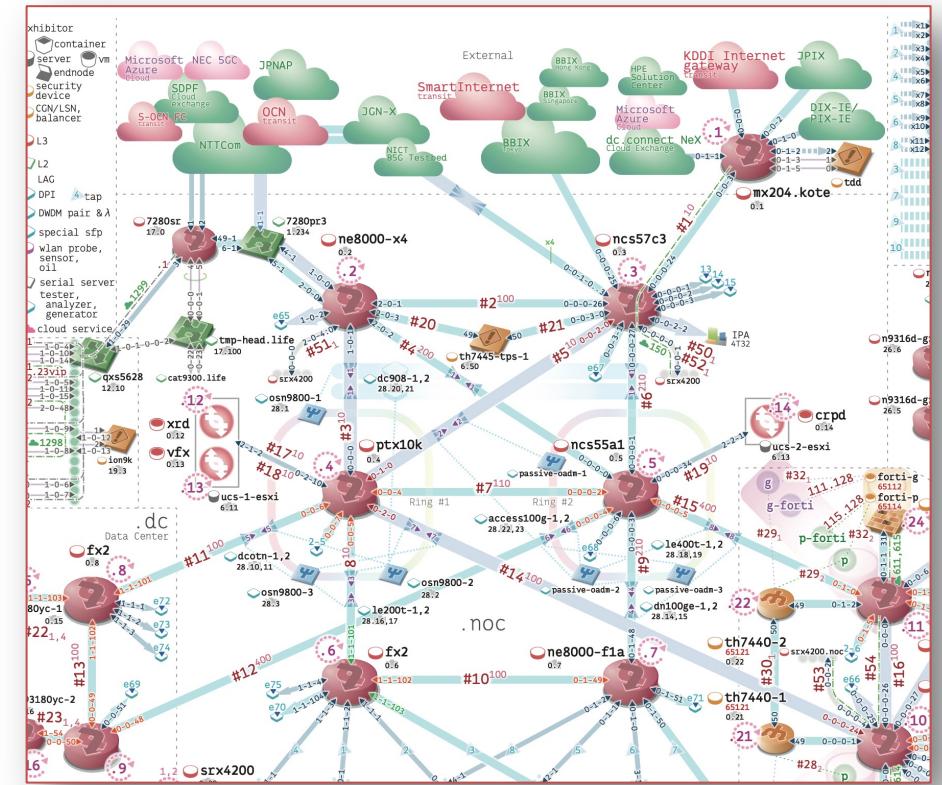
SRv6 SID in ShowNet2021

• SRv6 SID Design: 2001:3e8:6:deviceID:function::/64

SID	Behavior	Context	Owner	Status	MX
2001:3e8:6:1:1::	End (PSF)	'default':1	sidgr	Inuse	Y
2001:3e8:6:1:40::	End,DFE	'vrfl-global'	tgpr	Inuse	Y
2001:3e8:6:1:41::	End, X (PSF)	[way/W/1,254, Link-local]	1115:255	Inuse	Y
2001:3e8:6:1:42::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:256	Inuse	Y
2001:3e8:6:1:43::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:257	Inuse	Y
2001:3e8:6:1:44::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:258	Inuse	Y
2001:3e8:6:1:45::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:259	Inuse	Y
2001:3e8:6:1:46::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:260	Inuse	Y
2001:3e8:6:1:47::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:261	Inuse	Y
2001:3e8:6:1:48::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:262	Inuse	Y
2001:3e8:6:1:49::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:263	Inuse	Y
2001:3e8:6:1:50::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:264	Inuse	Y
2001:3e8:6:1:51::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:265	Inuse	Y
2001:3e8:6:1:52::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:266	Inuse	Y
2001:3e8:6:1:53::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:267	Inuse	Y
2001:3e8:6:1:54::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:268	Inuse	Y
2001:3e8:6:1:55::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:269	Inuse	Y
2001:3e8:6:1:56::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:270	Inuse	Y
2001:3e8:6:1:57::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:271	Inuse	Y
2001:3e8:6:1:58::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:272	Inuse	Y
2001:3e8:6:1:59::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:273	Inuse	Y
2001:3e8:6:1:60::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:274	Inuse	Y
2001:3e8:6:1:61::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:275	Inuse	Y
2001:3e8:6:1:62::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:276	Inuse	Y
2001:3e8:6:1:63::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:277	Inuse	Y
2001:3e8:6:1:64::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:278	Inuse	Y
2001:3e8:6:1:65::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:279	Inuse	Y
2001:3e8:6:1:66::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:280	Inuse	Y
2001:3e8:6:1:67::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:281	Inuse	Y
2001:3e8:6:1:68::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:282	Inuse	Y
2001:3e8:6:1:69::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:283	Inuse	Y
2001:3e8:6:1:70::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:284	Inuse	Y
2001:3e8:6:1:71::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:285	Inuse	Y
2001:3e8:6:1:72::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:286	Inuse	Y
2001:3e8:6:1:73::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:287	Inuse	Y
2001:3e8:6:1:74::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:288	Inuse	Y
2001:3e8:6:1:75::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:289	Inuse	Y
2001:3e8:6:1:76::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:290	Inuse	Y
2001:3e8:6:1:77::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:291	Inuse	Y
2001:3e8:6:1:78::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:292	Inuse	Y
2001:3e8:6:1:79::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:293	Inuse	Y
2001:3e8:6:1:80::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:294	Inuse	Y
2001:3e8:6:1:81::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:295	Inuse	Y
2001:3e8:6:1:82::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:296	Inuse	Y
2001:3e8:6:1:83::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:297	Inuse	Y
2001:3e8:6:1:84::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:298	Inuse	Y
2001:3e8:6:1:85::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:299	Inuse	Y
2001:3e8:6:1:86::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:300	Inuse	Y
2001:3e8:6:1:87::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:301	Inuse	Y
2001:3e8:6:1:88::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:302	Inuse	Y
2001:3e8:6:1:89::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:303	Inuse	Y
2001:3e8:6:1:90::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:304	Inuse	Y
2001:3e8:6:1:91::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:305	Inuse	Y
2001:3e8:6:1:92::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:306	Inuse	Y
2001:3e8:6:1:93::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:307	Inuse	Y
2001:3e8:6:1:94::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:308	Inuse	Y
2001:3e8:6:1:95::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:309	Inuse	Y
2001:3e8:6:1:96::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:310	Inuse	Y
2001:3e8:6:1:97::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:311	Inuse	Y
2001:3e8:6:1:98::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:312	Inuse	Y
2001:3e8:6:1:99::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:313	Inuse	Y
2001:3e8:6:1:100::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:314	Inuse	Y
2001:3e8:6:1:101::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:315	Inuse	Y
2001:3e8:6:1:102::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:316	Inuse	Y
2001:3e8:6:1:103::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:317	Inuse	Y
2001:3e8:6:1:104::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:318	Inuse	Y
2001:3e8:6:1:105::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:319	Inuse	Y
2001:3e8:6:1:106::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:320	Inuse	Y
2001:3e8:6:1:107::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:321	Inuse	Y
2001:3e8:6:1:108::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:322	Inuse	Y
2001:3e8:6:1:109::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:323	Inuse	Y
2001:3e8:6:1:110::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:324	Inuse	Y
2001:3e8:6:1:111::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:325	Inuse	Y
2001:3e8:6:1:112::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:326	Inuse	Y
2001:3e8:6:1:113::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:327	Inuse	Y
2001:3e8:6:1:114::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:328	Inuse	Y
2001:3e8:6:1:115::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:329	Inuse	Y
2001:3e8:6:1:116::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:330	Inuse	Y
2001:3e8:6:1:117::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:331	Inuse	Y
2001:3e8:6:1:118::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:332	Inuse	Y
2001:3e8:6:1:119::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:333	Inuse	Y
2001:3e8:6:1:120::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:334	Inuse	Y
2001:3e8:6:1:121::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:335	Inuse	Y
2001:3e8:6:1:122::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:336	Inuse	Y
2001:3e8:6:1:123::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:337	Inuse	Y
2001:3e8:6:1:124::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:338	Inuse	Y
2001:3e8:6:1:125::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:339	Inuse	Y
2001:3e8:6:1:126::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:340	Inuse	Y
2001:3e8:6:1:127::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:341	Inuse	Y
2001:3e8:6:1:128::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:342	Inuse	Y
2001:3e8:6:1:129::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:343	Inuse	Y
2001:3e8:6:1:130::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:344	Inuse	Y
2001:3e8:6:1:131::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:345	Inuse	Y
2001:3e8:6:1:132::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:346	Inuse	Y
2001:3e8:6:1:133::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:347	Inuse	Y
2001:3e8:6:1:134::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:348	Inuse	Y
2001:3e8:6:1:135::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:349	Inuse	Y
2001:3e8:6:1:136::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:350	Inuse	Y
2001:3e8:6:1:137::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:351	Inuse	Y
2001:3e8:6:1:138::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:352	Inuse	Y
2001:3e8:6:1:139::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:353	Inuse	Y
2001:3e8:6:1:140::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:354	Inuse	Y
2001:3e8:6:1:141::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:355	Inuse	Y
2001:3e8:6:1:142::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:356	Inuse	Y
2001:3e8:6:1:143::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:357	Inuse	Y
2001:3e8:6:1:144::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:358	Inuse	Y
2001:3e8:6:1:145::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:359	Inuse	Y
2001:3e8:6:1:146::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:360	Inuse	Y
2001:3e8:6:1:147::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:361	Inuse	Y
2001:3e8:6:1:148::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:362	Inuse	Y
2001:3e8:6:1:149::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:363	Inuse	Y
2001:3e8:6:1:150::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:364	Inuse	Y
2001:3e8:6:1:151::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:365	Inuse	Y
2001:3e8:6:1:152::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:366	Inuse	Y
2001:3e8:6:1:153::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:367	Inuse	Y
2001:3e8:6:1:154::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:368	Inuse	Y
2001:3e8:6:1:155::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:369	Inuse	Y
2001:3e8:6:1:156::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:370	Inuse	Y
2001:3e8:6:1:157::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:371	Inuse	Y
2001:3e8:6:1:158::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:372	Inuse	Y
2001:3e8:6:1:159::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:373	Inuse	Y
2001:3e8:6:1:160::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:374	Inuse	Y
2001:3e8:6:1:161::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:375	Inuse	Y
2001:3e8:6:1:162::	End, DFE (PSF)	[way/W/1,254, Link-local]	1115:376	Inuse	Y
2001:3e8:6:1:163					

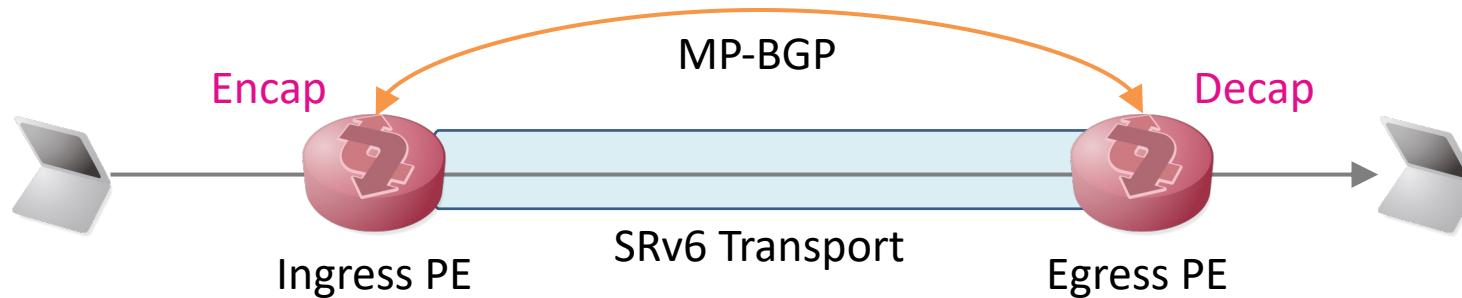
そして2022年

- ShowNetバックボーン全体がフルSRv6化
 - 4ベンダーから9機種のSRv6対応ルータと3種の仮想RR
 - Cisco、Huawei、Juniper、古河電工
 - BackboneはIPv6シングルルースタック
 - トラフィックは全てover SRv6で転送
- 相互接続検証および実験
 - SRv6 L3VPN、SRv6 Flex Algo、uSID
 - SRv6による長距離映像伝送実験



SRv6 Layer-3 VPN

- SRv6トランスポートのL3VPNという基本的なユースケース
 1. Ingress PEがH.EncapsでCEからのパケットをカプセル化
 2. PはただのIPv6パケットとして転送 or Traffic Engineering
 3. Egress PEはEnd.DT4/DT6でカプセル化を解除し転送
- MP-BGPによるPE間のL3VPNシグナリング
 - PE間でお互いの経路とそれに紐づくSIDを交換する
 - RFC9252で2022年7月に標準化完了(Interop Tokyo 2022よりも後)



SRv6 L3VPN 相互接続

- SIDと紐づくFunctionによってL3VPN/L2VPNが構築される
 - ルータ同士がBGPでSIDを交換しSRv6オーバーレイを構築
 - H.EncapsはPE内で完結、End.DTのSIDを対向PEへBGPで広告

```
RP/0/RP1/CPU0:ncs57c3.noc#show segment-routing srv6 sid  
Wed Jun 15 08:39:24.999 JST
```

```
*** Locator: 'shownet' ***
```

Locator	Type	Value	Manager	Status	Y/N
2001:3e8:fa00:3:1::	End (PSP/USD)	'default':1	sidmgr	InUse	Y
2001:3e8:fa00:3:40::	End.X (PSP/USD)	[Hu0/0/0/27.6, Link-Local]	isis-290	InUse	Y
2001:3e8:fa00:3:41::	End.X (PSP/USD)	[Hu0/0/0/26.2, Link-Local]	isis-290	InUse	Y
2001:3e8:fa00:3:42::	End.X (PSP/USD)	[Hu0/0/0/24.1, Link-Local]	isis-290	InUse	Y
2001:3e8:fa00:3:43::	End.DT6	'vrf-global'	bgp-290	InUse	Y
2001:3e8:fa00:3:44::	End.DT4	'vrf-global'	bgp-290	InUse	Y
2001:3e8:fa00:3:45::	End.DX2	310:310	l2vpn_srv6	InUse	Y
2001:3e8:fa00:3:46::	End.X (PSP/USD)	[FH0/0/2/0.5, Link-Local]	isis-290	InUse	Y

IPv6 Link Localアドレスの利用

- ・バックボーンのルータ間はIPv6 Link Localアドレスのみ
 - Link Localアドレスは自動生成される
➤ルータ間のリンクにはアドレスを設定する必要はない
 - descriptionとmtu、あとは機器によってipv6 enableやfamily isoなど
 - IGPで交換したLoopbackのGlobal UnicastアドレスでRRとBGP

```
interface FourHundredGigE0/0/0/24.14
  description fhg-0-2-0.ptx10k.noc
  mtu 9021
  ipv6 enable
```

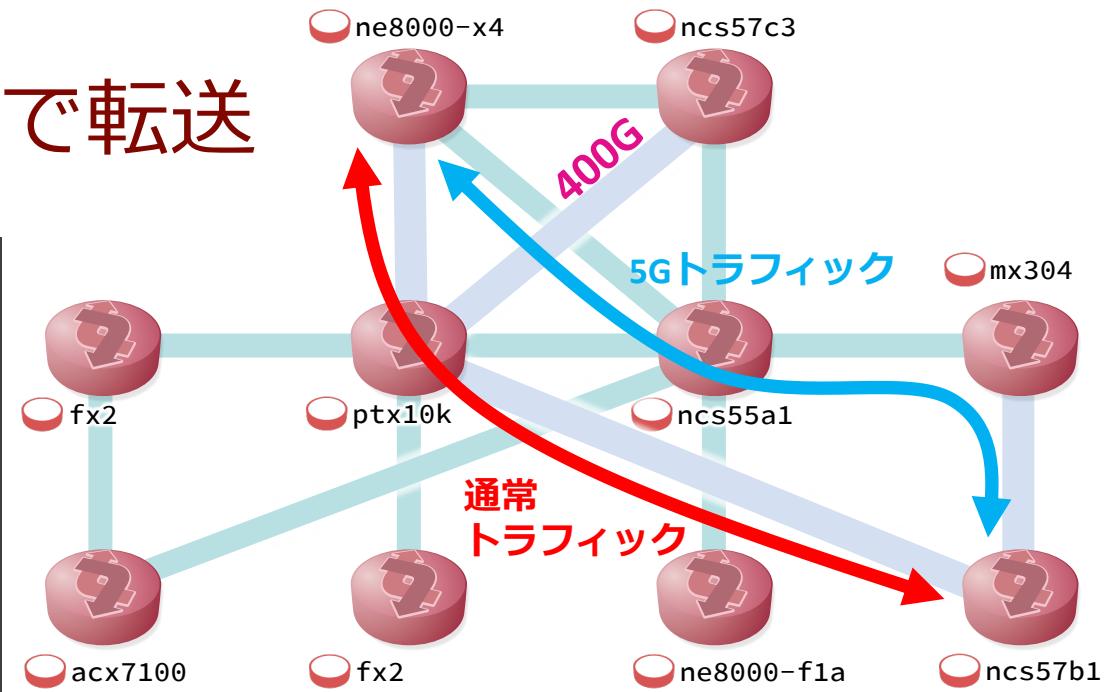
SRv6でのFlex Algo

- SRv6ではAlgoごとにLocatorをIGPに広告
 - 広告したAlgo(面)のSPF計算に含まれる
- Flex Algoで5Gのトラフィックを通常トラフィックとは別のパスで転送

```
mx304.svc# show protocols isis source-packet-routing
flex-algorithm [ 128 129 130 ];
srv6 {
    locator loc-mx304;
    locator loc-128;
    locator loc-129; ↗ それぞれ別のIPv6 Prefix (Locator)
    locator loc-130;
}
```



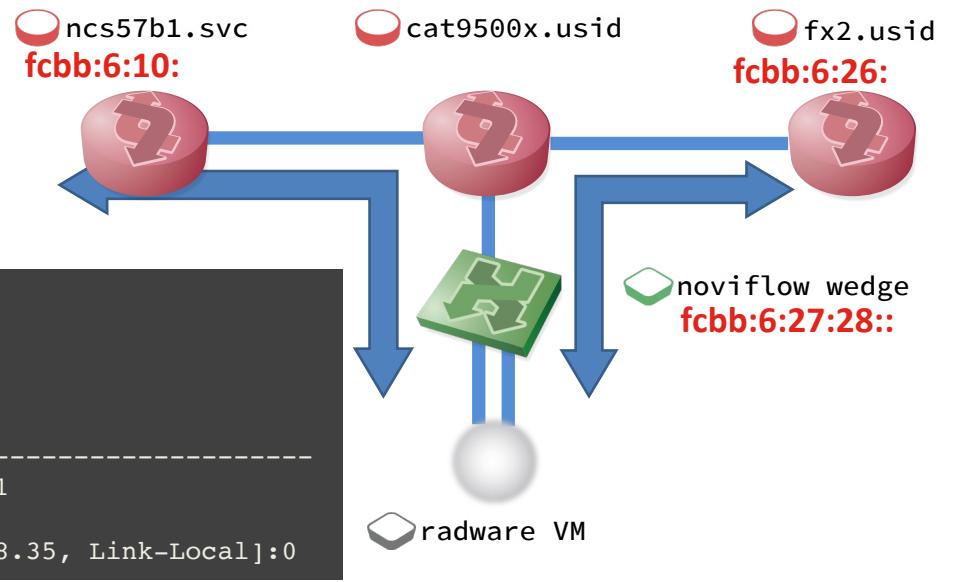
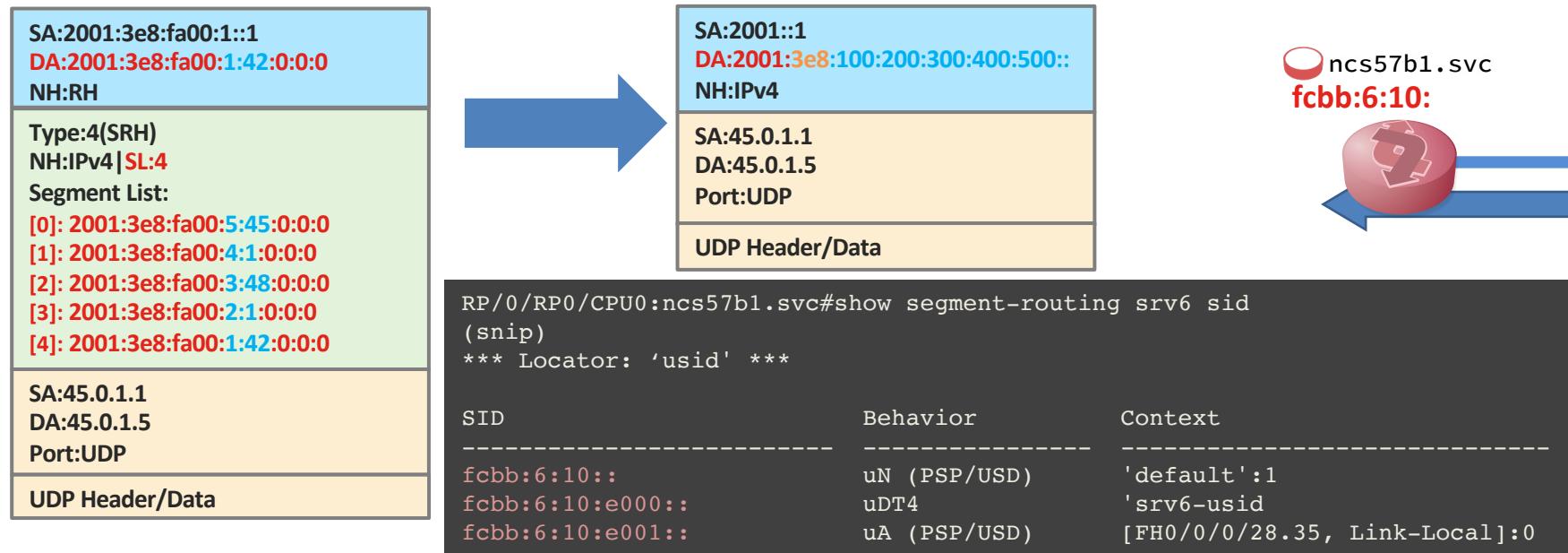
通常トラフィックが通っていた
リンクは400G-ZR+



SRv6 uSID Interoperability Test

draft-filsfils-spring-net-pgm-extension-srv6-usid

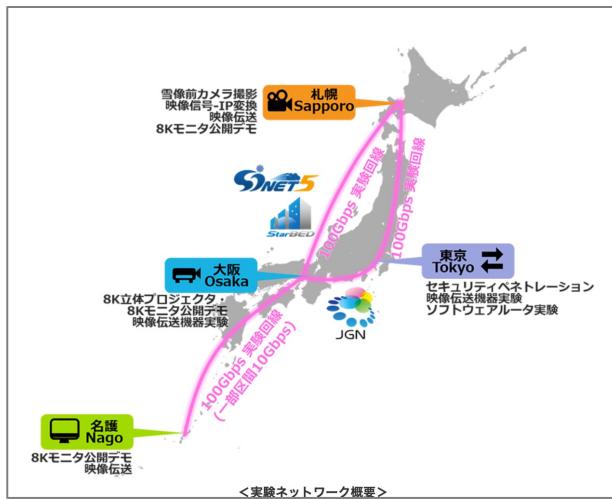
- SRv6 uSIDの相互接続試験を実施
 - NCS57b1とFX2間でF3216 formatのuSIDをISIS/BGPにて広告、疎通を確認
 - ExplicitにPath制御することでNoviflow社SWのuSIDなService Chaining Function(End.SC)の通せることを確認



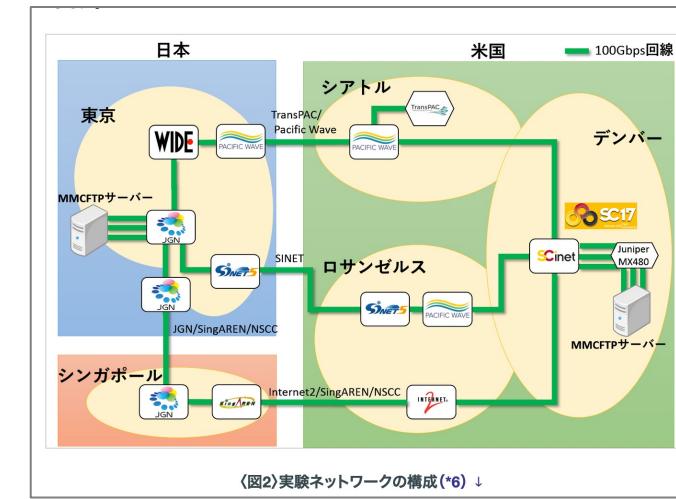
SRv6による長距離伝送路の構築実験

・背景: 長距離大容量伝送実験

- ・本当に性能が出るかはやってみないとわからないことが多い
- ・一方で、時には海外に渡る長距離伝送路を確保するのは大変な労力(調整/構築/運用)を伴う



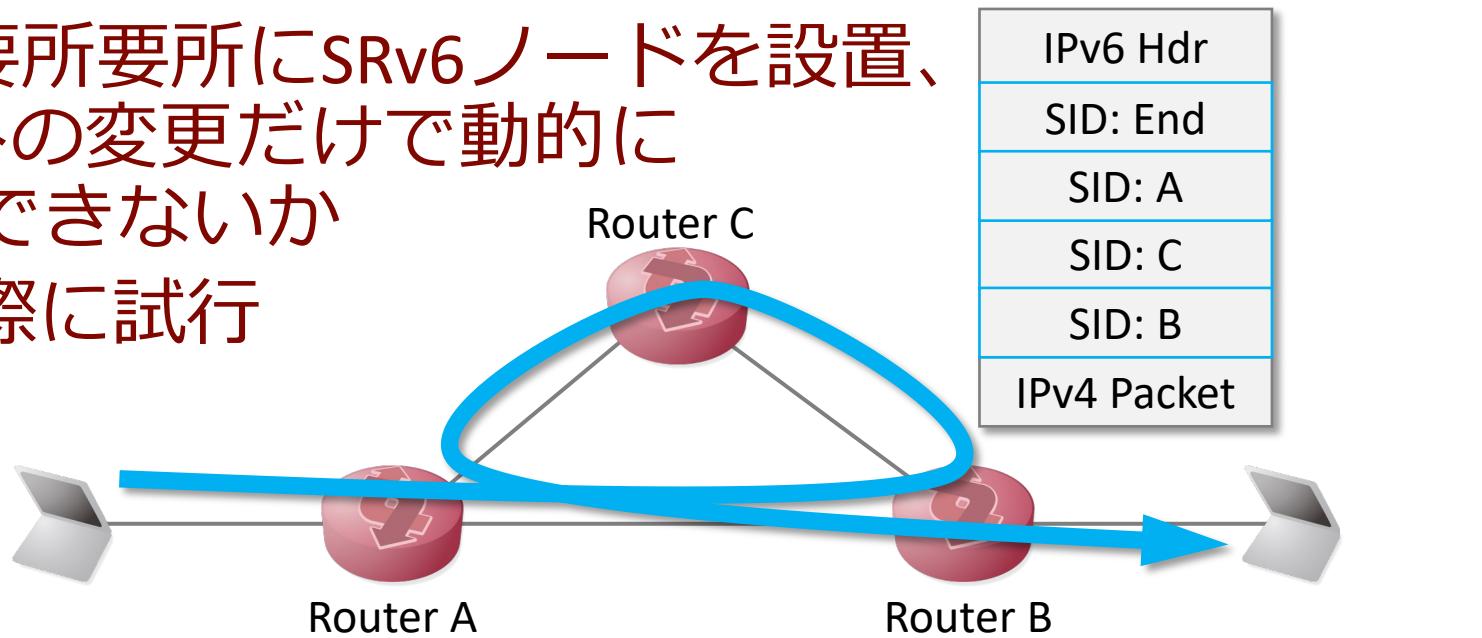
NICT, 2本の非圧縮フル解像度8K映像による超高精細立体映像伝送に成功,
<https://testbed.nict.go.jp/event/yukimatsuri2020-press.html>



NII, ファイル転送プロトコルMMCFPTで転送速度231Gbpsを達成／長距離データ転送の世界記録を更新, <https://www.nii.ac.jp/news/release/2017/1214.html>

SRv6を使った実験

- SRv6のIPv6透過性を活かした仮想/動的な長距離伝送路
 - SRv6であれば、Headendで挿入するSegmentリスト次第で、パケットに同じパスを往復させることも可能
 - IPv6ネットワークの要所要所にSRv6ノードを設置、あとはSegmentリストの変更だけで動的に長距離伝送路を構築できないか
 - ShowNetとWIDEで実際に試行



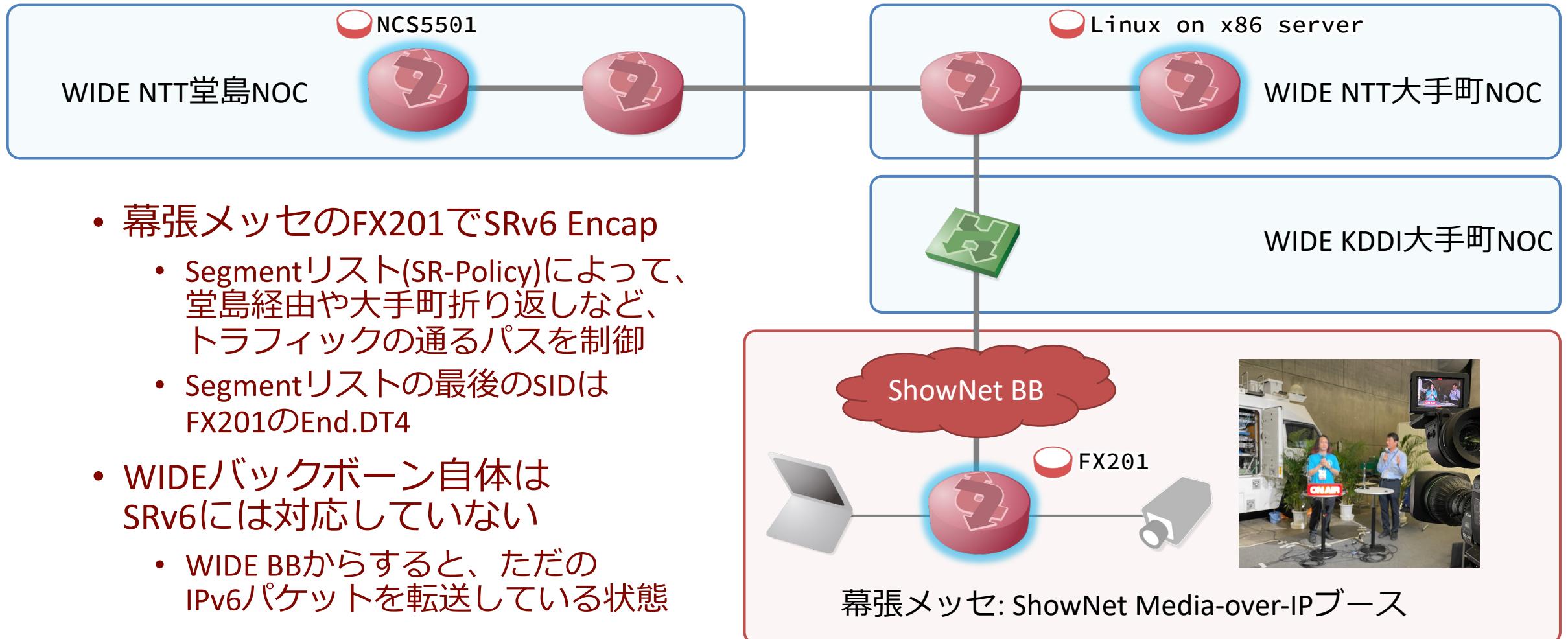
SRv6による長距離映像伝送を実施

- ・東京と大阪にSRv6ノードを設置
- ・ShowNetのMedia-over-IPのデモ環境から SMPTE-ST2110のストリームをSRv6 Encap、 東阪を往復して戻して幕張メッセで再生
- ・展示期間中ブースの 映像を継続して転送

ShowNet構築期間中に
ラック前でテスト映像が
無事再生された際の様子

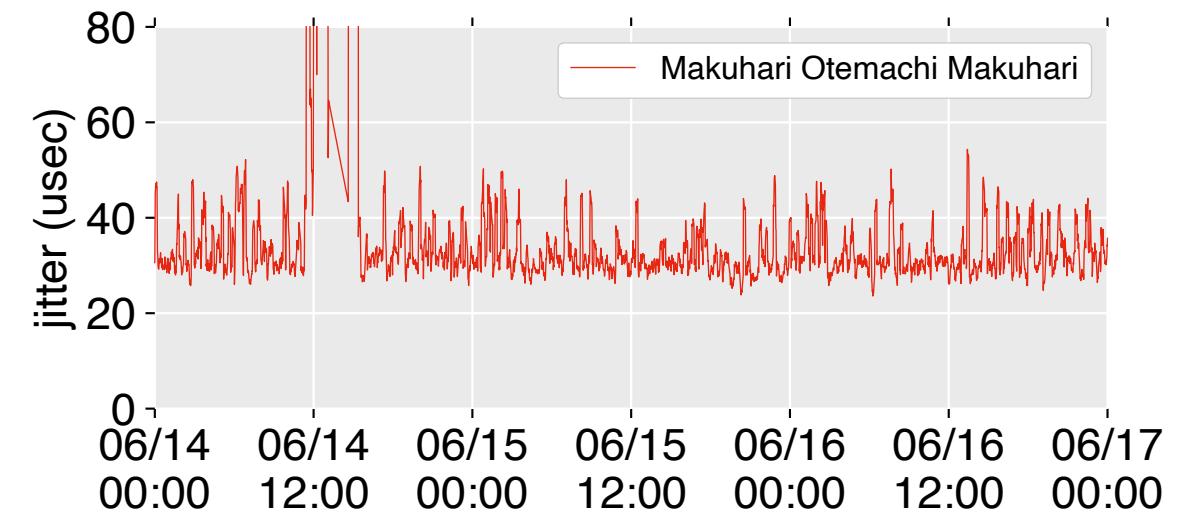
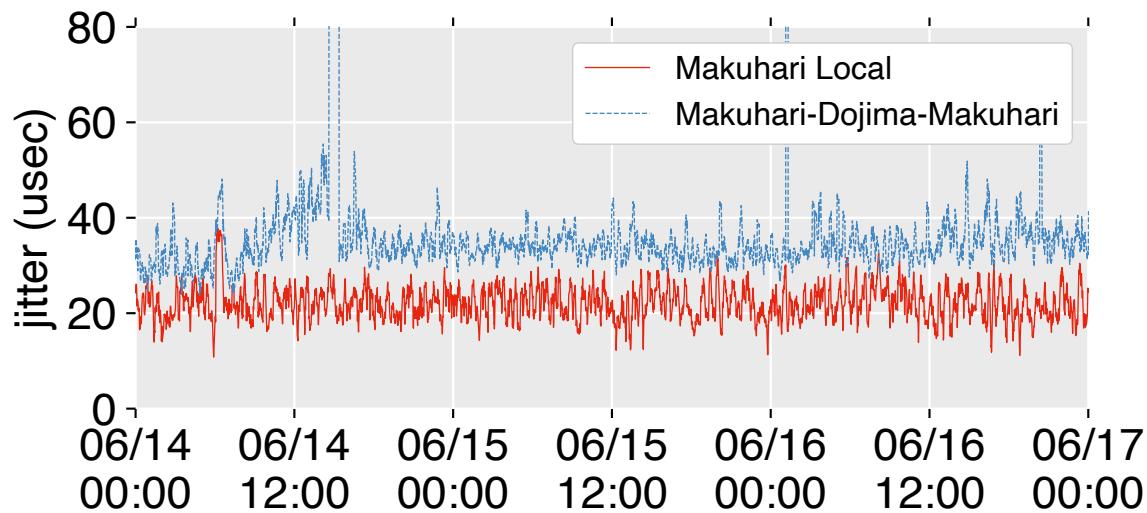


実際の構成



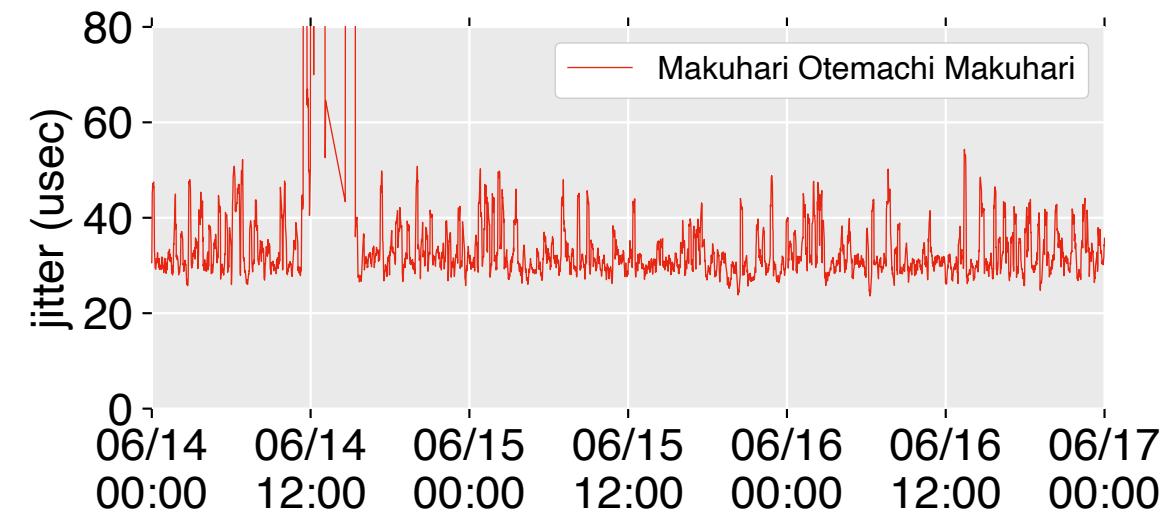
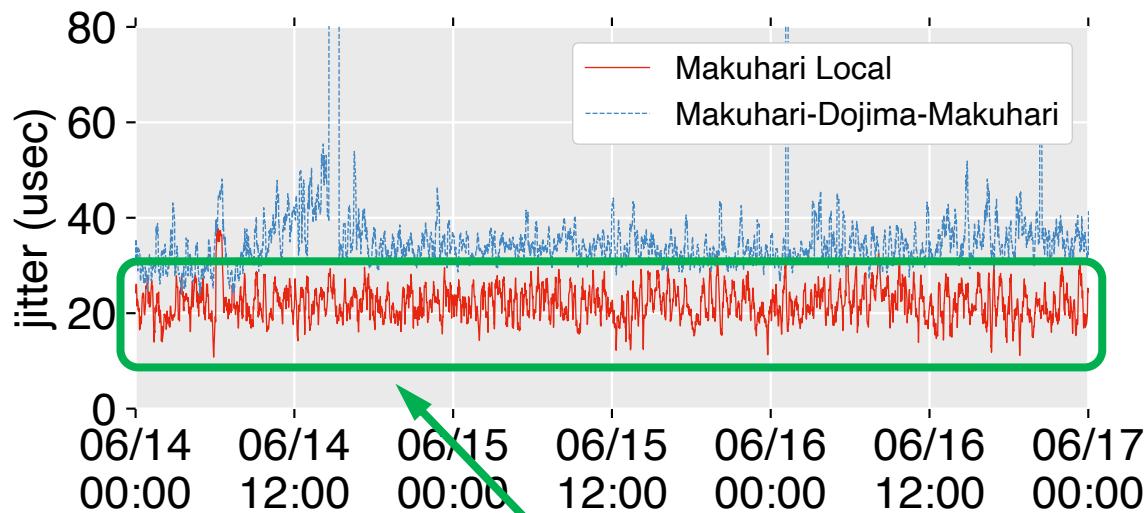
ジッターの計測

- 映像伝送とともに、伝送路のジッター計測も実施



ジッターの計測

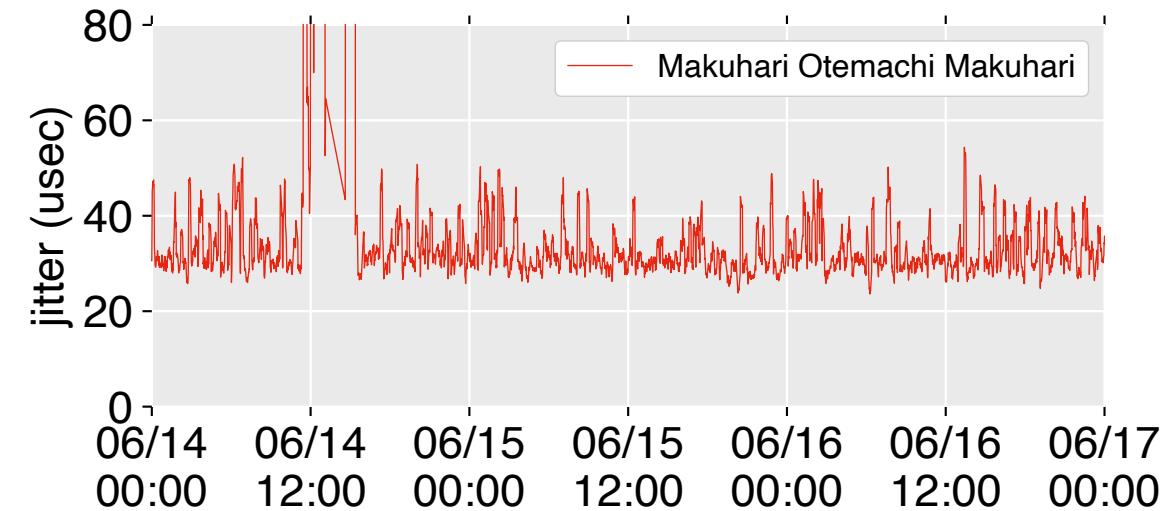
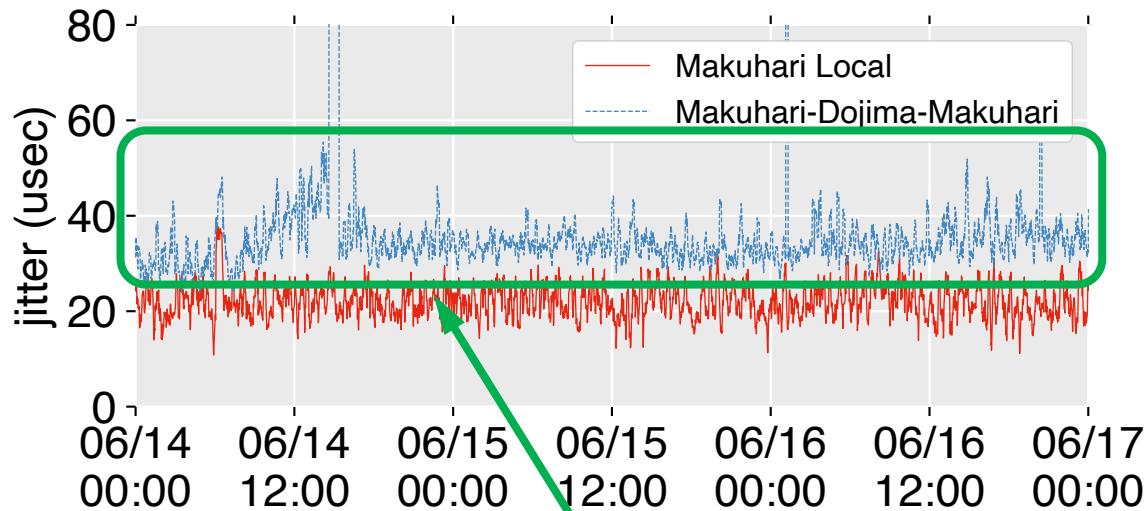
- 映像伝送とともに、伝送路のジッター計測も実施



ローカル環境での折り返しがもっともジッターが小さく20usecほど。
参考: SMPTE ST2110-30(AES67)が1Kpps = 1ms Packet Interval Time (PIT)

ジッターの計測

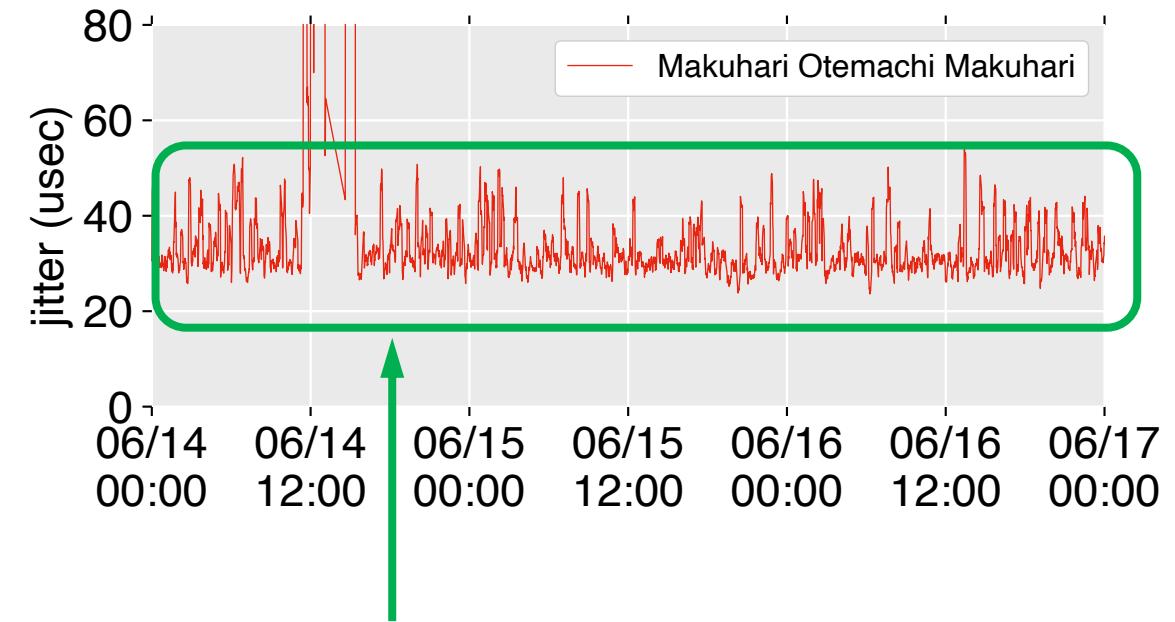
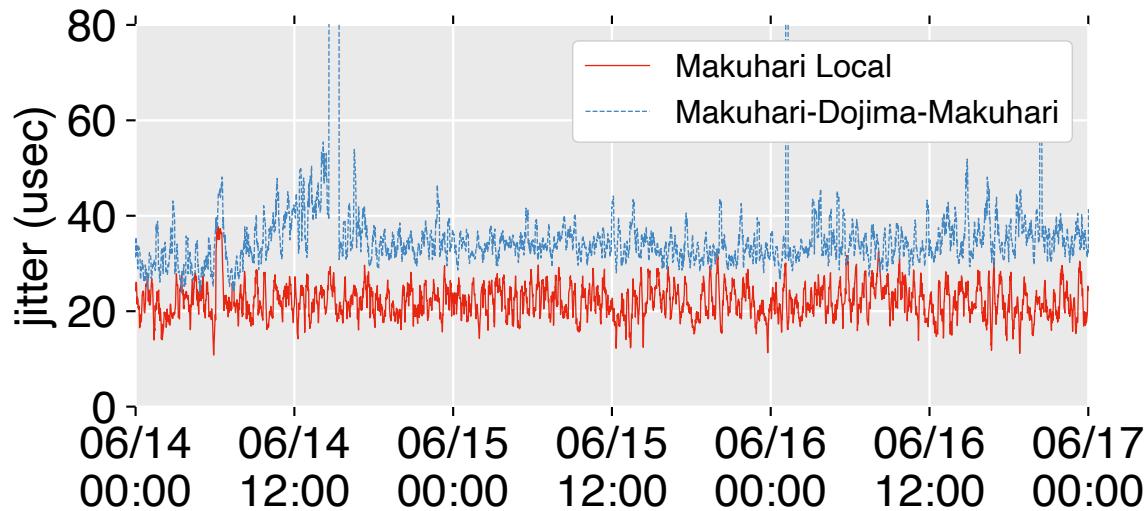
- 映像伝送とともに、伝送路のジッター計測も実施



幕張→堂島→幕張と東阪を往復すると30~40 usecと大きくなるが許容範囲。
WIDEバックボーンの途中経路が輻輳していなかつたことは確認

ジッターの計測

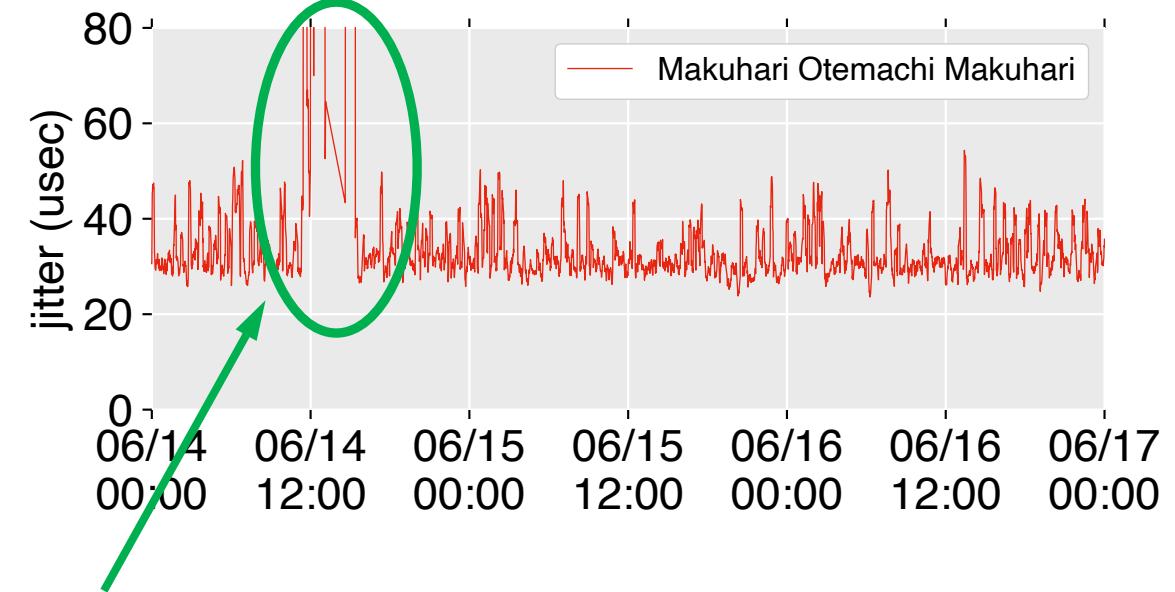
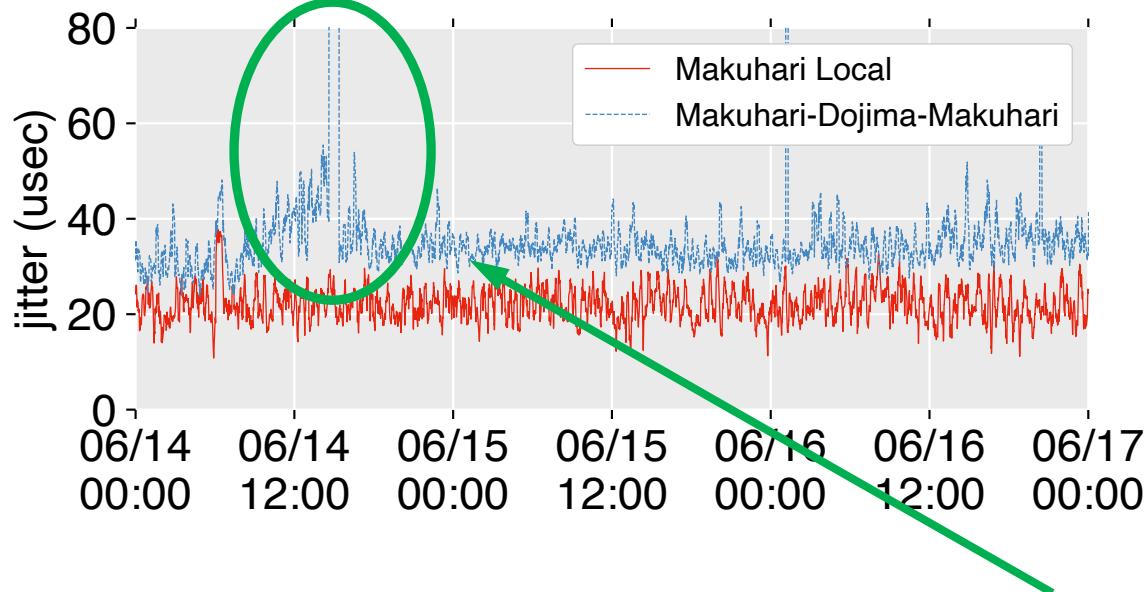
- 映像伝送とともに、伝送路のジッター計測も実施



一方、幕張→大手町→幕張では、ジッターは堂島往復よりは小さい。
x86サーバのLinux kernelのソフトウェア転送のジッターはそれほど大きくはない様子

ジッターの計測

- 映像伝送とともに、伝送路のジッター計測も実施



Peerが上がっていく過程で意図せずShowNetとWIDEの間に
別のASが入ってしまい、急激にジッターが悪化

Lessons Learned

- SRv6でASを跨いだ動的な長距離伝送路の構築はできる
 - IPv6ネットワークの要所要所にSRv6ノードを設置するだけ
 - あとはHeadendルータで挿入するSegmentリストでパスを制御
- SRv6 Encapは $40 + 8 + \text{SID数} \times 16\text{-byte}$ のMTUオーバーヘッド
 - 3箇所通過と96-byte、ST 2110の最大パケットサイズは1460-byte
 - そのため今回はjumbo frameの通るWIDE経由のみで実験を行なった
 - uSIDなどで緩和は可能
- トラフィックが意図しないネットワークを通過する事故
 - 今回はjumboの通るネットワークだったためスループットのみ落ちた
 - SRv6-awareなOAM機能(e.g., traceroute)がないと気付くのが難しい

Lessons Learned cont'd

- SRv6ヘッダ付きのパケットを透過的に転送しない機器
 - WIDEバックボーン中で大幅なスループットの減少を確認
 - 対応していないIPv6拡張ヘッダ(SRH)のパケットをCPUに上げるルータが存在した
- パケットに埋め込むSIDの数に関する制約
 - 機器によって処理できる1パケット中のSIDの段数が異なる
 - MTUが続く限りパスを構成するようなことはできない

まとめ

- ShowNetにおけるSRv6 Single Stack Backbone
 - 4ベンダー9機種によるSRv6 L3VPNの相互接続
 - IPv6 Link Localアドレスのみでバックボーンを構築可能
 - Flex-AlgoなどのTE機能もSRv6で相互接続しつつ動作
- SRv6による長距離映像伝送路の構築実験
 - 構築自体は少ない手間でできることを確認
 - 一方でMTUやOAM、SIDの段数やIPv6透過性に関する課題も
 - 今回は準備期間が短かったが、より緻密な実験/計測ができる可能性も



ちょっと待ってほしい

プログラマビリティ...? 🤔

- RFC8986 SRv6 Network Programming
 - *The concept of "SRv6 Network Programming" refers to the capability of an application to encode any complex program as a set of individual functions distributed through the network.*
- 2022年にできたのはL3VPNだけ
 - SDNやNFVでできたほどのプログラマビリティ(または柔軟性)は、まだ発揮できてはいない(と、個人的には思う)

プログラマビリティとの悪戦苦闘史: 導入編

- 2012年~2013年: OpenFlowの導入
- ネットワークオペレータがプログラミングもし始める
 - ネットワークの運用とプログラミングは以前は別個のものだった
 - ネットワーク側にもプログラミングできる"余地"が生まれ始める
- プログラミングする目的
 - 新しい機能/付加価値を実現する手段
 - 自動化



SDN Japan 2014

#show△net ←
Scratch and Re-build the Internet

ネットワーク運用自動化の難しさ

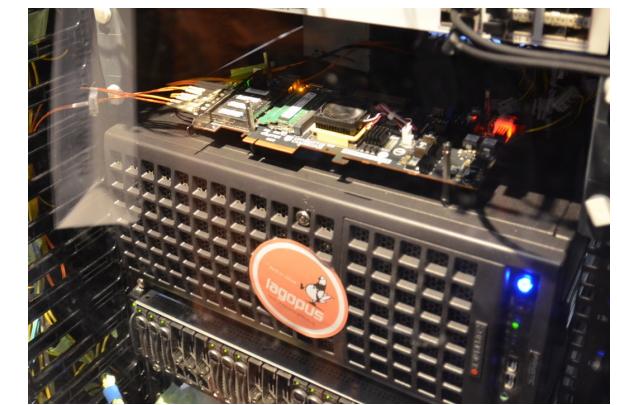
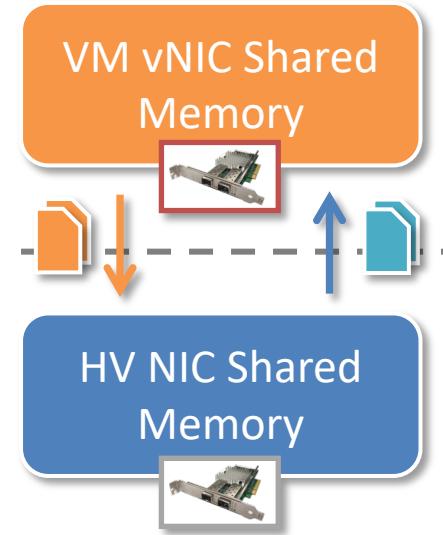
- ソフトウェア制御で設定の変更を自動化
- ベンダーや製品、バージョンごとに異なるAPI
 - 実際の現場ではさまざま機器を利用している
 - 全ての機器を網羅するAPIは現状では存在しない
- エラーハンドリングの難しさ
 - Debuaggabilityの欠如

Copyright © INTEROP TOKYO 2014 ShowNet NOC Team

57

プログラマビリティとの悪戦苦闘史: 仮想化編

- 2014年~2016年: 仮想ルータの登場
- ソフトウェアによるパケット転送が現実的に
 - 仮想ルータ自体をソフトウェア的にインテグレーションする
 - ハードウェアのコモディティ化
- 性能が課題に
 - ネットワークオペレータがコンピューターキテクチャも勉強し始める
 - 時にはFPGAにも手を出す



むき出しのNetFPGA-SUMEと、
LagopusとVirNOSの動作するx86サーバ

プログラマビリティとの悪戦苦闘史: 回帰編

- 2017年~2018年: IPでもがんばれるのでは?
- IP Routingでサービスチェイニングを実現する
 - VRFとBGP Flowspecの合わせ技
 - Flowspec Routeを打ち込むコントローラをexabgpで実装
 - FlowspecをFlow Modのように使ったとも言える



振り返って：プログラマできる余地の拡大

- ・プログラマビリティは様々なレイヤーに存在する
 - ・厳密には、プログラムから触れるAPIが各層に整備された
 - ・下に行くほど大変だが、新しいことを実現できる
 - ・上に行くほど(一見)簡単なインターフェースで制御を書ける

Orchestration	ONOS, Vendor Orchestrator, 必要な部分だけ自分で作ったり
Management	YANG, Netconf, SNMP, gNMI, gNOI, Ansible, etc
Forwarding Rule	Routing Protocols, PCEP, Flowspec, Netlink, Cisco XR Service Layer, Juniper JET, etc
Switch Chip	P4, SAI, Vendor SDK, etc

抽象化とその漏れ

- ・プロトコル/ネットワークスタックとは、抽象化そのもの
 - ・ある層はその下の層を隠蔽し、抽象化された機能を実現する
 - ・例: TCPは下のIPv4やIPv6を隠蔽し、信頼性のある通信を提供する
 - ・例: 機器のOSはASICを隠蔽し、ルーティングプロトコルによってネットワークにおけるノードとしてふるまうことを実現する
 - ・現実には、抽象化は漏れる
 - ・下のレイヤーのことを知らないと、上のレイヤーの問題を解決できない



図 2 現実問題では抽象化は漏れる(不完全になる)。

なぜソフトウェア論文を書くのは難しい(と感じる)のか,
権藤克彦 et.al., コンピュータソフトウェア, 2009
原典はLaw of Leaky Abstractions, Joel Spolsky

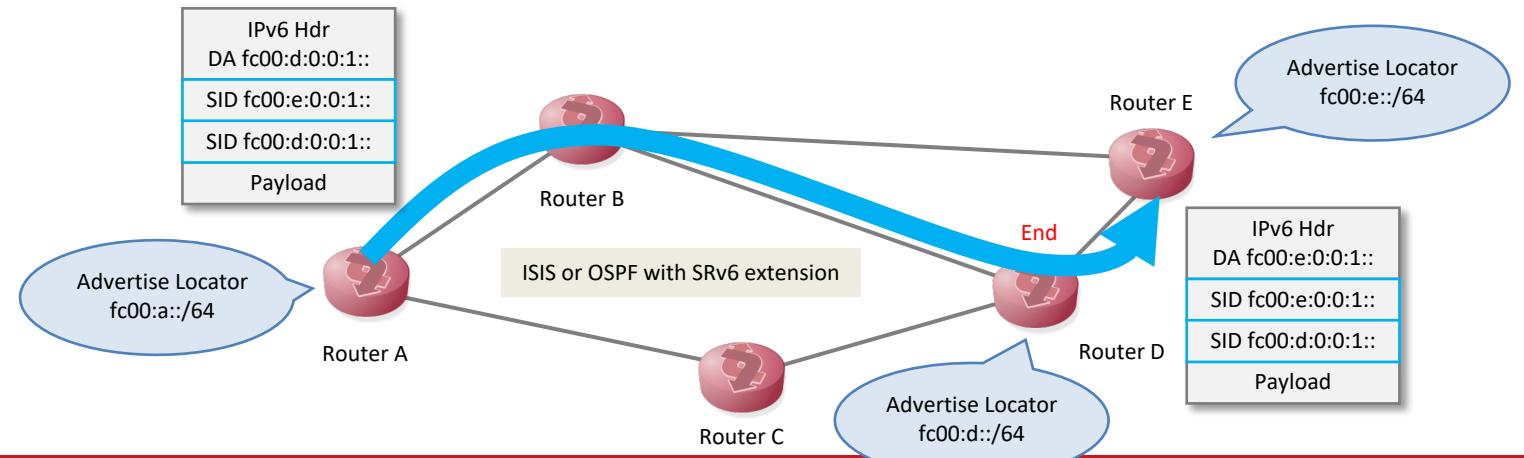
抽象化と漏れとデバッグと

- 抽象化の漏れ@ネットワーク
 - 例: 壊れているのは、自分が書いたOpenFlowアプリか？コントローラか？Secure Channelか？はたまたOpenFlowスイッチか？
 - オペレータならおそらくみんな似たような経験しているのでは
- よく出来た抽象化とよく出来たその実装、つまりプロトコル、ソフトウェアを見出す/生み出すことが大事
- ShowNetでの悪戦苦闘は、この部分の成熟度合いに起因(早すぎたんだ...)



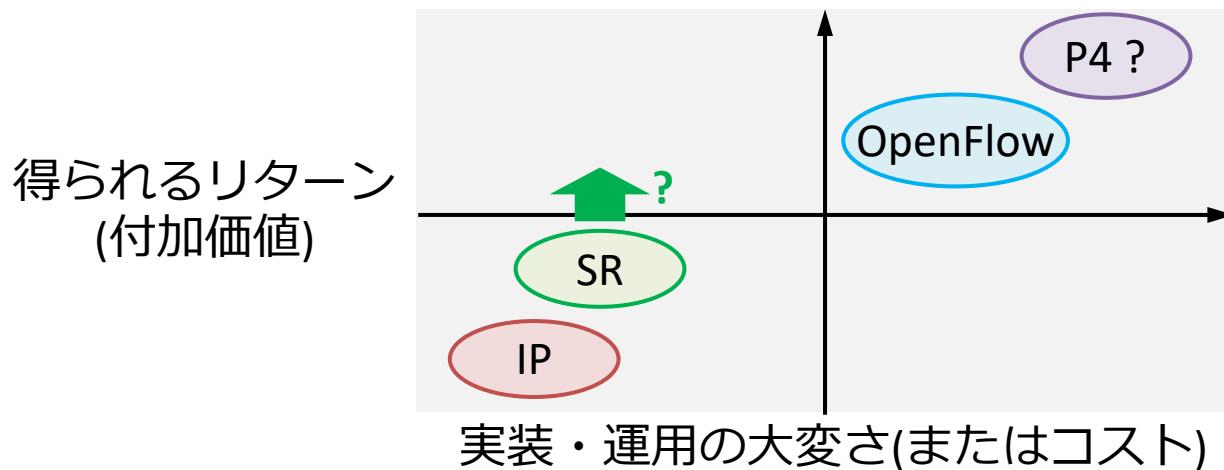
Segmentという抽象化

- Source Routingという機能と、実装としてのSR(v6)
 - その下のレイヤーは信頼と実績のプロトコルと機器群
 - 今ついにSRv6というレイヤーが動くところまで来た
 - 次はこのレイヤーの上に何を作るのか
 - SRv6でプログラミングする余地はどこだ？



悪戦苦闘の果てに

- 柔軟なネットワークの実現に向けて@Interop Tokyo ShowNet
 - ネットワークがプログラマビリティを獲得する10年
 - OpenFlowはrawすぎて、それだけを現場で使うにはハードルが高かった
 - ソフトウェアエコシステムのネットワークへの導入が始まったSDN/NFV
 - 実践的な運用の楽さと、柔軟性を実現する複雑さのバランスをとったソリューションとしてのSource Routingの再発見



おわりに

- 何のプログラマビリティをどう活かすか
 - プログラマビリティは様々なレイヤーに存在する
 - 開発・運用の大変さと、得られるリターンのバランス
 - SRv6には引き続き期待
- 来年もぜひShowNetに足をお運びいただければ幸いです
 - 先んじて悪先苦闘
 - そしてコミュニティへの還元



Interop®22 Tokyo

JUNE 15-17th
MAKUHARI MESSE, JAPAN

SHOWNET

Over the Premise

