

Multipath Transport Protocol を用いた SRv6 BGP Egress Peer Engineering による高品質・高信頼なコンテンツ配信

豊田 安信^{1,2,a)} 三島 航^{3,b)} 早坂 彪流^{4,c)} 深川 祐太^{3,d)} 植原 啓介^{5,e)} 中村 修^{5,f)}

受付日 2023年11月19日, 採録日 2024年4月15日

概要: コンテンツ事業者 (CSP) のサービス品質改善には, CSP からユーザに至るいくつかの経路のうち, より通信品質の良いものを用いることが求められる. BGP で対外接続を行う CSP では, (1) BGP に品質を示すパラメータが無い点や, (2) BGP 経路 (Egress Path) の性能評価と選択ポリシーの随時適用が難しい点, そして, (3) コンテンツによって要求品質が異なる点から, 品質に基づいた経路選択・利用が困難であった. 本研究では, これらの課題を解決し, 品質に基づいて Egress Path を利用する制御機構を「EPE コントロールプレーン」と定義し, その実現手法として “Multipath Transport SRv6-EPE Suite” を提案する. 本手法では通信品質の最適化に MPTCP を, Egress Path の共有と利用に SRv6 L3VPN を活用する. 本手法を用いることでアプリケーションごとの要求品質に合わせたリアルタイムな経路選択・制御を実現できる. 仮想環境での評価実験では, 本手法は BGP ベストパスを利用した通常の TCP による従来のコンテンツ配信と比較して, 一部のテストケースでは品質を大幅に向上させることが明らかとなった. 特に通信品質の変化を想定したテストケースでは, 通信経路上に障害が発生した際にもサービスの断絶なく配信可能であることを確認した. また, 国内外の主要な ISP を通じた実環境での評価実験では, 9 つの ISP 中 6 つで 10% 以上スループットが向上し, うち 4 つの ISP では 50% 以上の顕著な向上がみられた.

キーワード: egress peer engineering, BGP, MPTCP, SRv6, トラフィックエンジニアリング

Performance Aware Content Delivery Using Multipath Transport Protocol with SRv6 BGP Egress Peer Engineering

YASUNOBU TOYOTA^{1,2,a)} WATARU MISHIMA^{3,b)} TAKERU HAYASAKA^{4,c)} YUTA FUKAGAWA^{3,d)} KEISUKE UEHARA^{5,e)}
OSAMU NAKAMURA^{5,f)}

Received: November 19, 2023, Accepted: April 15, 2024

Abstract: To improve service quality of Content Service Providers (CSPs), it is necessary to select a better path from CSPs to users. It is difficult for CSPs using BGP for external connections to deliver content based on path quality because: (1) there are no BGP attributes for indicating quality, (2) there are no general methods for the performance evaluation of BGP routes (egress paths) and dynamic applying of BGP Policy, and (3) quality requirements are vary depending on the content. To address these issues, we define a control mechanism for utilizing egress paths based on quality as the “EPE Control Plane” and present the “Multipath Transport SRv6-EPE Suite” as a method for its realization. This approach leverages MPTCP for optimizing communication quality and SRv6 L3VPN for sharing egress paths. This enables real-time path selection tailored to the different requirements depending on the applications. In the virtual environment evaluations, our method significantly improved throughput in certain test cases compared to regular TCP content delivery with the BGP best path, especially in scenarios anticipating communication quality changes, ensuring uninterrupted service even when network failures occur. Moreover, real-world evaluations through major domestic and international ISPs showed throughput improvements of over 10% in 6 out of 9 ISPs, with 4 ISPs experiencing substantial increases of more than 50%.

Keywords: egress peer engineering, BGP, MPTCP, SRv6, traffic engineering

1. はじめに

インターネットを介しコンテンツを提供する事業者 (Content Service Provider: CSP) のサービスにおいては、ネットワークの通信品質 (Quality of Service: QoS) がユーザの満足度に大きく影響することが知られている [1]. 多くの CSP は、冗長性や収容容量の拡大を目的として複数のトランジット事業者や Internet Exchange (IX) と接続し、複数の対外接続経路 (Egress Path) を持つ。たとえば Facebook の自律システム (Autonomous System: AS) は、クライアントに対して平均 3 つの Egress Path を有すると報告されている [2]. CSP はこれらの経路から Border Gateway Protocol (BGP) [3] のベストパス選択ルールにより 1 つの Egress Path を利用し、コンテンツを配送する。

BGP ベストパスは必ずしも通信品質に即して選択されるわけではなく、しばしばコンテンツ配信のパフォーマンス劣化を引き起こすことが知られている [4], [5]. 下記の主な 3 つの原因から、BGP によって品質に基づいた経路選択を行うことは困難である。

(1) BGP は直接的に品質を示す経路のパラメータを持たない

現行の BGP 経路決定メカニズムは、経由する AS の数を表す AS_PATH 属性や、運用者が適宜行う重み付け (BGP ポリシー) に依存しているため、直接的に Egress Path のパフォーマンスに基づいた経路選択ができない。QoS メトリクスを用いた経路交換の試みとして、BGP の拡張属性 [6] や独自の Software Defined Network (SDN) 機構を利用したもの [7] が提案されている。しかし、これらの手法は多くの AS で広く導入されるまでメリットを受けにくい課題があるほか、AS の競争力に関わるネットワーク内部の品質情報を他の AS に公開する必要があるため、現在に至るまでインターネット全体で広く利用されているものは

存在しない。

(2) Egress Path の性能評価と選択ポリシーの随時適用が困難

顧客までの通信経路上には、CSP の管理下に無いいくつかの事業者の AS のネットワークが存在する。そのため実際にトラフィックを流して計測しなければ、経路品質を評価することはできない。仮に通信品質が良い経路を発見できたとしても、通信品質は流動的であり、トラブルの原因となり得る品質低下は常に観測可能とは限らないため、解決がより困難になっている。実際に、コンテンツサービスの配送品質に関する問題が発生した場合の原因調査と特別な BGP ポリシーの管理には、多くの労力と時間が必要となる [8].

(3) コンテンツによって要求品質が異なる

コンテンツ利用者が体感する品質 (Quality of Experience: QoE) は、本質的に定量化が困難な概念である。しかし、測定可能な QoS メトリクスが顧客の QoE に及ぼす影響を定量的に評価するため、一定の仮定のもとに QoE 値を推定する取り組みが様々な領域で行われている [9]. 当然ながら、コンテンツの種別によって推定 QoE 値を適切に導出するためのモデル [10]) は異なる。そのため、AS 全体で宛先ネットワークごとに原則として 1 つのパスを選択する (BGP ベストパスのみを利用する場合、いくつかのコンテンツサービスでは最適な Egress Path を利用できないことになる。

つまり、CSP が多様な隣接 AS と接続し、利用できる Egress Path の数を増加させたとしても、常にユーザやプロバイダが希望するパフォーマンスを得られるかどうかは未知数である。

BGP のベストパス選択ルールによらず Egress Path を選択し、アウトバウンドトラフィックを制御する技術は Egress Peer Engineering (EPE) と呼ばれる [11]. EPE を用いた経路選択により、コンテンツ配信品質の向上が見込めることが、プロダクション環境を利用した実証実験から明らかになっている [12]. 以降本稿では、エンドツーエンド (E2E) でのパフォーマンス向上を目的とした EPE を特に “Performance Aware EPE” と呼称する。

高品質な BGP パスを利用して継続的にサービスを提供するためには、利用可能な Egress Path を適切に管理し、コンテンツサービスに最適なものを選択・利用する必要がある。本研究では、Performance Aware EPE の実現のために求められる役割と機能を整理し、SRv6 L3VPN と Multipath TCP (MPTCP) を活用した EPE 用のコントロールプレーンスタック “Multipath Transport SRv6-EPE Suite” を提案する。このコントロールプレーンは、他の AS に協力を仰ぐことなくコンテンツ配信品質の向上を見込める点や、通信品質や BGP 経路の変化に動的に追

¹ 慶應義塾大学政策・メディア研究科
Graduate School of Media and Governance, Keio University,
Fujisawa, Kanagawa 252-0882, Japan

² 株式会社ブロードバンドタワー
BroadBand Tower, Inc., Chiyoda, Tokyo 100-0011, Japan

³ エヌ・ティ・ティ・コミュニケーションズ株式会社
NTT Communications Corporation, Chiyoda, Tokyo 100-8019,
Japan

⁴ BBSakura Networks 株式会社
BBSakura Networks, Shinjuku, Tokyo 160-0023, Japan

⁵ 慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University,
Fujisawa, Kanagawa 252-0882, Japan

a) yas-nyan@sfc.wide.ad.jp

b) w.mishima@ntt.com

c) t-hayasaka@bbsakura.net

d) y.fukagawa@ntt.com

e) kei@wide.ad.jp

f) osamu@wide.ad.jp

従できる点、そして、多様なコンテンツサービスに適合する柔軟性がある点で特徴的である。

はじめに、インターネットコンテンツ配信環境を模した仮想環境での実験を通じて、Multipath Transport SRv6-EPE Suite の有用性を評価した。すべてのテストケースにおいて、本提案手法は従来の BGP ベストパスのみを利用したコンテンツ配信と同等以上の性能を有したほか、いくつかのケースにおいては配信品質を大きく向上できることを明らかにした。また、ネットワーク環境の変化を想定した2つのテストケースにおいても、本提案手法は通信経路上の品質劣化のコンテンツ提供品質に対する影響を大きく低減できることが分かったほか、BGP 経路の変更にも迅速に対応可能なことを示した。

次に、実際のインターネット AS 上で動作するコンテンツサーバに Multipath Transport SRv6-EPE Suite を実装し、日本の主要な4社のモバイル ISP を含む9つの ISP に接続するサンプルクライアントに対して、実際にコンテンツサービスを提供する評価実験を行った。この評価実験では、BGP ベストパスのみを利用した従来の配信環境と比較して、9つ中6つの ISP で10%以上スループットが向上し、うち4つの ISP では50%以上の顕著なスループット向上を確認できた。

本稿の貢献は以下のとおりである。

- EPE コントロールプレーン：ネットワークの状況に応じて動的にトラフィックを制御する (Performance Aware EPE) ために必要な機構の設計
- “Multipath Transport SRv6-EPE Suite” の提案
- 実際のコンテンツ配信環境での有用性の評価と、今後の技術課題の整理

2. 要件定義

2.1 想定するネットワーク

本研究が前提とする、CSP AS が IPv6 インターネットを介して様々なコンテンツ配信を行うネットワークの概略図を図1に示す。CSP は自社のネットワークを AS として運用し、内部にはコンテンツサーバや外部接続を担うルーター (Autonomous System Boundary Router: ASBR) などが存在する。CSP は Transit AS, Public Peer, Private Network Interconnect (PNI) など、複数の外部接続点 (Egress Peer) を持ち、BGP ベストパス選択ルールに基づき、宛先ネットワークごとに一意の Egress Path を選択する。図1の例では、3つの Egress Path の中から Egress Path α を選択して利用している。

また、CSP は多様なコンテンツを提供する故にそれぞれスループットや遅延、パケットロスなどの QoS に対して異なる要求基準を持つ。しかし、通信経路上には CSP の管理下に無いネットワークも存在するため、コンテンツの提供に際し、これらの基準を事前に満たすことは困難で

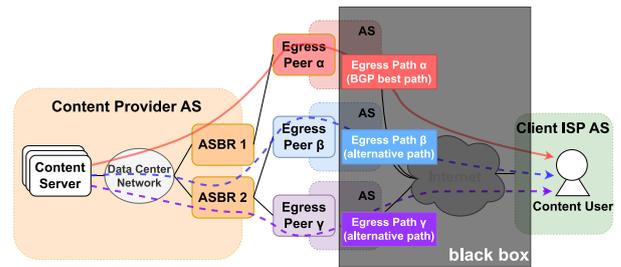


図1 CSP と利用者間のネットワーク

Fig. 1 Network between the Content Service Provider (CSP) and content clients.

ある。

CSP は複数のデータセンターにまたがり構成される場合もあるが、全拠点を同一のポリシーで運用する際はルーティング上の差異が存在しないため、本稿では考慮しない。

2.2 SRv6-EPE

Segment Routing (SR) は、ソースルーティングに基づくトラフィックエンジニアリングの実現手法である [13]。パケットの転送対象を Segment と呼ばれる単位で扱い、SR Header (SRH) に埋め込んだ Segment List を用いて、パケットに与える転送処理を制御する。SR 技術のうち、IPv6 拡張ヘッダを利用するものに SRv6 [14] がある。

SRv6 のセマンティクスでは、IPv6 アドレス形式で記述する Segment Identifier (SID) と呼ばれる識別子を用いて Segment を表現する。SID は、そのパケットを經由させるノードを表す “Locator” と、そのノードで行う処理 (SRv6 Behavior) について記述する領域によって構成される。SRv6 はこの SID の表現性を利用して様々なネットワークサービスを活用できる。

本研究では Segment Routing による EPE 運用モデル [11] のうち、SRv6 の仕組みを利用した EPE 実現手法を “SRv6-EPE” と呼称する。SRv6-EPE は、ASBR がパケットを Egress Peer ごとにユニークな SID (以後 “SRv6-EPE SID” と呼ぶ) で識別し、送信元ホストが SRv6-EPE SID を付与することで、BGP ベストパスによらずにパケットを転送できる。この仕組みにより、アプリケーションレベルでの Egress Path 選択が可能となり、既存の BGP ベストパス選択の制限を超えた自由なトラフィック配信が実現される。

SRv6-EPE は他の EPE 実現手法と比較して、(1) 中継ネットワークに非依存である点、(2) EPE を利用しないトラフィックと共存できる点、(3) 既存の IP ルーティングの技術を活用できる点で優れているため [12]、本研究では SRv6-EPE により EPE を実現する環境を想定する。

2.3 Performance Aware EPE

本研究では、BGP ベストパスにかかわらずパフォーマンスに基づいて Egress Path を選択し、通信品質向上を目指す取り組みを“Performance Aware EPE”と呼ぶ。第1章で述べたように、オペレータがコンテンツに適した Egress Path を手動で選択し、随時これをサービスに適用するのは困難である。そのため、Performance Aware EPE を実現するためには、ネットワーク状況の変化に応じて動的にトラフィックを制御する機構が必要になる。この機構には下記の2点の働きが求められる。

- (1) CSP AS が利用可能な Egress Path の把握と共有
- (2) Egress Path のパフォーマンス評価とポリシーの随時適用

(1) の実現には SRv6-EPE を適切に動作させるネットワーク環境の整備が、(2) の実現にはコンテンツの配信品質を最大化する Egress Path の選択と利用を行う仕組みがそれぞれ求められる。本稿ではこれらを兼ね備えた制御機構を「EPE コントロールプレーン」と定義し、必要となる機能を分類して整理する。

2.4 EPE コントロールプレーンに求められる要求事項

第2.1節で定義された CSP のサービスでは、変動するネットワーク環境上でコンテンツごとに異なる品質要件を満たす必要があるため、EPE コントロールプレーンには以下の要件が求められる。

- (1) コンテンツ配信品質の向上

Performance Aware EPE は、通信品質が優れた Egress Path を適切に選択し、配信品質を向上させることを目的とする。実世界では CSP の管理外の AS が経路上に存在するため、実際にトラフィックを送出することによる End-to-End でのパフォーマンス測定と評価が必要である。適切に Egress Path を評価するためには、実際に利用者に提供されるコンテンツの通信品質を指標とするのが望ましい [12]。

- (2) 通信環境の変化への追従性

第1章で述べたように、インターネットの通信品質や到達性は日常的に変動するため、EPE コントロールプレーンは、これらの変化に対応し、コンテンツサービスの中断や品質悪化を防ぐ必要がある。

- (3) 多様なコンテンツへの柔軟性

第1章でも述べたように、コンテンツサービスの種類や特徴によって求められる品質基準は異なる。EPE コントロールプレーンには、これら多様な QoS 品質要件に対応する柔軟性が求められる。

- (4) 展開の容易性

第1章で触れたように、現状の BGP には直接的に品質を示すパラメータがなく、品質情報を AS 間で交換するための機能拡張も、多くの AS に導入されるまではメリットを享受することが難しい。EPE コントロールプレーンは、隣接する協力無しに CSP AS に展開可能なことが求められる。また、新規機能を既存サービスに追加する際には、サービスの継続性を維持することが課題となる [15]。CSP が EPE コントロールプレーンを導入する際には、既存のネットワークやサービスへの影響を最小限に抑えつつ、ビジネスの継続性を保つことが必要である。さらに、展開に必要な新たな機構を最小限に抑えることは、運用に掛かる負荷の軽減につながる。

3. EPE コントロールプレーン

本章では、EPE コントロールプレーンに必要となる機能群を定義し、それぞれの目的と担う役割について述べる。

図2はEPE コントロールプレーンの各機能およびデータプレーンとの相互関係を示したものである。EPE コントロールプレーンには、その性質から“EPE Provisioning and Reachability Management”と“Quality-Based Egress Path Decision Making and Traffic Control”の2つの役割

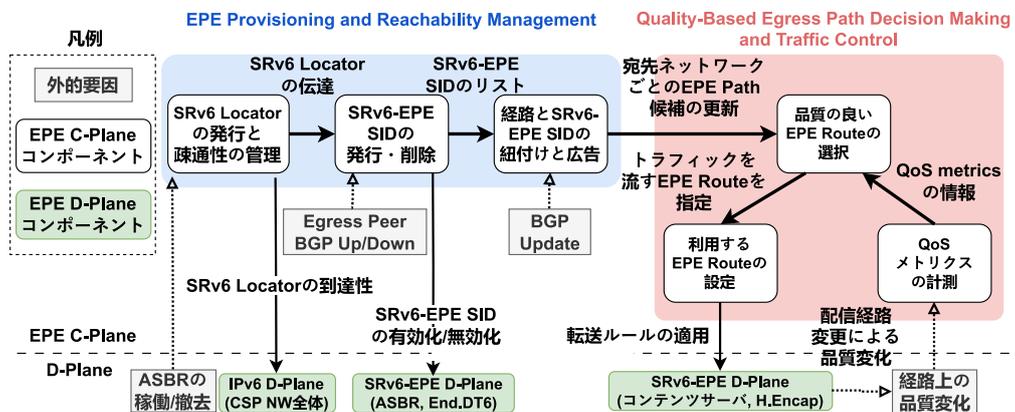


図2 EPE コントロールプレーンの各コンポーネントと相互作用

Fig. 2 Components and interactions in the EPE control plane.

が存在する。

3.1 EPE Provisioning and Reachability Management

第 2.4 節で述べた「SRv6-EPE を適切に動作させるネットワーク環境の整備」を担う役割を、本研究では“EPE Provisioning and Reachability Management”と定義する。これは下記の機能から構成される。

- **SRv6 Locator の発行と到達性の管理**

SRv6 SID は Locator と Function および Argument で構成される [14]。SRv6-EPE において Locator は ASBR ごとにユニークに割り当てられる prefix であり、CSP ネットワーク全体からの IPv6 到達性が必要となる。これはネットワーク内ですでに運用されている Interior Gateway Protocol (IGP) の経路制御機構によって実現される。

- **SRv6-EPE SID の発行・削除**

ASBR は自身の Locator と各 Egress Peer ごとにユニークな識別子を組み合わせる SRv6-EPE SID を生成し、その SID が付与されたパケットを SRv6 ヘッダを除去したうえで、対応する Egress Peer に転送するために必要なルールをルーティングテーブルに設定する。Egress Peer との接続や BGP 隣接関係が失われた際にルールを削除し、不要なトラフィックの送出を防ぐ機能も必要となる。

- **経路と SRv6-EPE SID の紐づけと広告**

CSP の Egress Peer にはインターネット全体への疎通性を提供するトランジット事業者のほかに、自身の顧客の経路のみを広告する ISP 事業者や、一部の地域の経路のみを提供する“Partial Transit”事業者 [16] が存在しうる。そのため、すべての Egress Peer から等しく任意の宛先に到達可能であるとは限らない。Egress Peer から広告されていない経路宛の通信をその Egress Peer に送出する行為は、その通信が宛先まで到達しない可能性があるのみならず、Egress Peer のネットワーク容量を窃取する不当な行為と見なされる可能性がある [17]。実際に中村らが 2021 年に行った調査 [18] では、多くの AS がこのような行為に無防備であったことが報告されている。BGP の経路選択によらずにトラフィックを制御する EPE を運用する場合、意図的なものであれ偶発的なものであれ、Egress Path の適切な利用は最も懸念すべき事項である。

したがって、Egress Peer から広告される経路は、特定の IP 経路情報と SRv6-EPE SID を関連付けて管理されなければならない。この組み合わせを本稿では“EPE Route”と呼称する。SRv6-EPE の対象トラフィックを生成するコンテンツサーバには、サービス提供前に必要な

EPE Route を伝達する必要がある。また、Egress Peer から広告が停止した経路に関連する EPE Route は速やかにホストから削除することも求められる。

3.2 Quality-Based Egress Path Decision Making and Traffic Control

EPE コントロールプレーンの機能のうち、第 2.4 節で述べた「コンテンツの配信品質を最大化する Egress Path の選択と利用」にあたる役割を“Quality-Based Egress Path Decision Making and Traffic Control”と定義する。

第 2.4 節で述べたように、Egress Path を適切に評価するためにはトラフィックを実際に送る必要がある。下記の 3 つの機能が相互に作用することで、パフォーマンスがより優れた EPE Route を利用したコンテンツ配信が可能となる。

- **QoS メトリクスの計測**

ある EPE Route を利用した際のコンテンツ配信のパフォーマンスを、各種 QoS メトリクスを収集・蓄積することで計測する。収集するメトリクスはコンテンツの種類や用途に合わせて設定される。

- **品質の良い EPE Route の選択**

収集された QoS メトリクスを参照し、EPE Provisioning and Reachability Management から得られた複数の EPE Route から、コンテンツサービスの QoS 基準に基づき、最も適切な EPE Route を判断する。

- **利用する EPE Route の設定**

選択された EPE Route を実際にコンテンツトラフィックが利用するための準備処理を行う。たとえば新たな転送ルールをデータプレーンに設定したり、パケットが通る EPE Route を直接指定する処理が想定される。これにより、クライアントへのコンテンツの配信経路が変更される。

4. 提案手法

本研究では、第 3 章で述べた EPE コントロールプレーンの実現手法として“Multipath Transport SRv6-EPE Suite”を提案する。Multipath Transport SRv6-EPE Suite は大きく分けて 2 つの要素で構成される。1 つ目は“EPE Provisioning and Reachability Management”を担う SRv6 L3VPN、2 つ目は“Quality-Based Egress Path Decision Making and Traffic Control”を担う MPTCP である。本章ではこれらの技術の概要と具体的な活用手法を述べる。

4.1 SRv6 L3VPN による EPE Provisioning and Reachability Management

4.1.1 SRv6 L3VPN

SRv6 L3VPN は、コントロールプレーンに Multi Protocol BGP (MP-BGP)、データプレーンに SRv6 を使用する

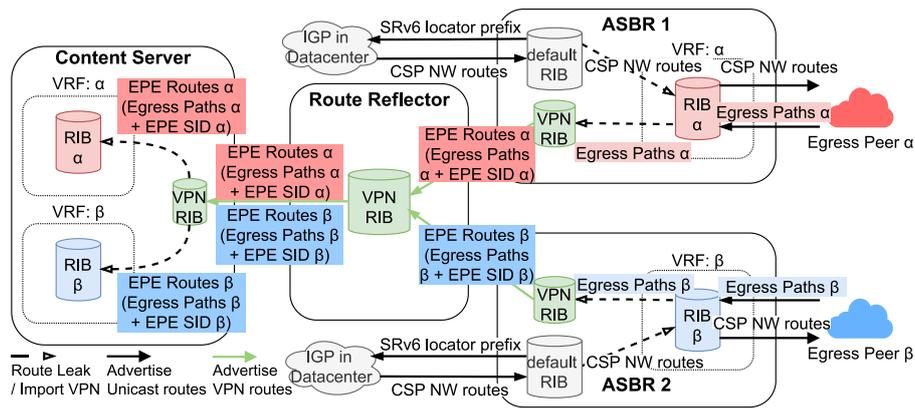


図 3 Multipath Transport SRv6-EPE Suite における経路交換
 Fig. 3 Route exchange in the Multipath Transport SRv6-EPE Suite.

る L3 オーバーレイサービスであり [19], IPv6 VPN 経路 [20] を通じて複数テナントの IPv6 疎通性を提供する。パケットは各 IPv6 VPN 経路に付随する SRv6 SID を用いて、ルータ上の適切な Virtual Routing and Forwarding (VRF) に転送される。この転送には SRv6 Behavior の End.DT6 または End.DT46 が利用される [14]。SRv6 L3VPN を有効化したルータでは、ルーティングデーモンが VRF への転送に必要な SRv6 SID の発行とフォワーディングルールの作成が自動的に行われる。したがって、第 3.1 節で述べた “EPE Provisioning and Reachability Management” の機能のうち、「SRv6-EPE SID の発行・削除」は ASBR 上のルーティングデーモンが行うことになる。

4.1.2 EPE に利用する経路の取り扱い

Multipath Transport SRv6-EPE Suite では、クライアントとのパケットの送受信に際し、非対称的な経路制御を行うことで SRv6-EPE を実現する。

図 3 は、図 1 のネットワークで SRv6 L3VPN を用いた SRv6-EPE を行うために必要な経路情報のやり取りを、Egress Peer α と Egress Peer α に関連する部分を抜粋し表したものである。

Performance Aware EPE を行うためには、各 Egress Peer から受信した Egress Path を区別して保持する必要がある。Multipath Transport SRv6-EPE Suite は、Egress Peer ごとに個別の VRF をサーバに用意することで Egress Path を区別する。ASBR からが Egress Peer から受信した経路は SRv6-EPE SID を含む IPv6 VPN 経路 (EPE Route) として、ルータリフレクタを経由してコンテンツサーバに配布され、それぞれの VRF にインストールされる。この一連の経路共有プロセスは、“EPE Provisioning and Reachability Management” の役割の一部として、ASBR、ルータリフレクタ、コンテンツサーバのルーティングデーモンが分散的に協調動作して行われる。

一方で、サーバがクライアントから受信するパケットに

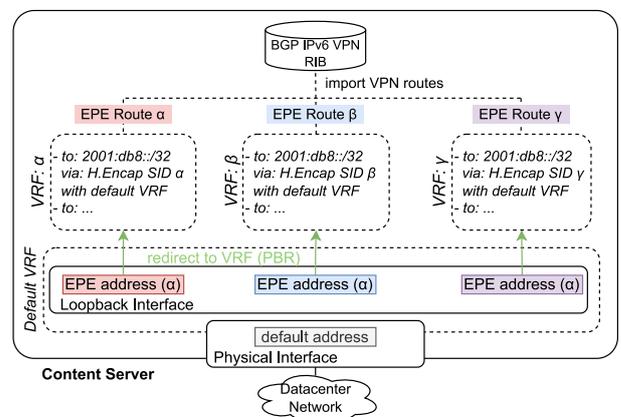


図 4 コンテンツサーバ内の VRF と EPE 専用アドレスの関係
 Fig. 4 Relationship between VRFs and EPE-specific addresses in a content servers.

は、CSP から Egress Peer に広告される通常の IPv6 経路が使用される。ASBR の各 Egress Peer 接続インタフェースは専用の VRF に属しているため、受信した CSP ネットワークの経路はこの VRF にルートリーク^{*1}される。

第 3.1 節で述べたように、SRv6-EPE SID の上位ビットは ASBR を示す SRv6 Locator プレフィックスが占めている。“EPE Provisioning and Reachability Management” の構成機能の 1 つである「SRv6 Locator の発行と到達性の管理」を充足するためには、ASBR を示す SRv6 Locator プレフィックスへの経路を、通常の IPv6 経路として CSP ネットワーク内に広告する必要がある。

4.1.3 Policy Based Routing による Egress Path の利用

コンテンツサーバには通常のインタフェースアドレス (以後デフォルトアドレス) に加えて、Egress Peer ごとに特殊な IP アドレス (以後 EPE 専用アドレス) を設定す

^{*1} ここでは出力先の VRF でその経路を利用したフォワーディングできるようにすることを指す。

る。EPE 専用アドレスによるアウトバウンド通信には、対応する Egress Peer から受け取った Egress Path のみが明示的に利用される。図 4 は、図 1 のネットワーク上のコンテンツサーバの、EPE 専用アドレスの利用手法を示したものである。サーバは Egress Peer α, β, γ から受信した経路^{*2}が EPE Route としてそれぞれ保持される VRF と、3 つの EPE 専用アドレスが属する通常のルーティングテーブル（デフォルト VRF）を持つ。EPE 専用アドレスを使用する通信は、デフォルト VRF に設定された Policy Based Routing (PBR) により、対応する VRF にリダイレクトされる。

4.2 Multipath TCP による Quality-Based Egress Path Decision Making and Traffic Control

4.2.1 MPTCP

MPTCP は、複数の TCP コネクションを統合して 1 つのセッションとして利用する TCP 拡張であり [21]、複数のアクセス回線の帯域の有効活用や、疎通性が不安定な環境下での冗長性の確保を目的とした適用が進んでいる [22]-[25]。MPTCP が動作するホストは、MPTCP の通信に利用可能なインタフェースアドレスを、“MPTCP Endpoint” として管理する。通常は異なるアクセス経路ごとに Endpoint が設定される。MPTCP セッションを確立したホスト間で Endpoint の情報を交換することにより、セッションに参加する TCP コネクション (Subflow) の確立と管理が行われる [26]。

4.2.2 MPTCP セッション確立までの流れ

図 5 は、図 1 の CSP 上のサーバが Multipath Transport SRv6-EPE Suite を用いてコンテンツサービスを提供する際の、トラフィックの流れを簡易的に示したものである。本提案手法ではコンテンツサーバとクライアントとの間で MPTCP セッションを確立し、トラフィックを送出する Subflow をパケットごとに選択することで、コンテンツの配信品質の最大化を目指す。

第 4.1 節でも述べたようにコンテンツサーバは二種類のアドレス（デフォルトアドレス、EPE 専用アドレス）を有する。デフォルトアドレスはその時点で BGP ベストパ

スとなる Egress Path を、EPE 専用アドレスは対応する Egress Peer から受信した Egress Path を常に利用して、アウトバウンド通信が行われる。なお、EPE 専用アドレスが利用する Egress Path は、第 4.1 節で記した仕組みにより事前にコンテンツサーバにインストールされる。

本提案手法ではこれらのアドレスそれぞれで Subflow を確立する。たとえば図 1 のネットワークでは、CSP は α, β, γ の 3 つの Egress Peer を有するため、BGP ベストパスを常に利用する Subflow と、Egress Path α, β, γ を明示的に利用する 3 つの Subflow の、合計 4 つの Subflow が利用されることになる。

本提案手法における各 Subflow が確立されるまでの手続きについて述べる。はじめに、クライアントはコンテンツサーバのデフォルトアドレスに対して TCP コネクション確立を要求する。この TCP コネクション（以後初期コネクションと呼ぶ）が確立された後、クライアントが MPTCP をサポートしていれば、MPTCP セッションが確立される。この初期コネクションは以後 Subflow の 1 つとなり、他の Subflow が確立されるまでのコンテンツ配信を担う。なお、第 3.1 節で述べたように、すべての Egress Path から任意のクライアントまで必ずしも到達可能であるとは限らない。そのためコンテンツサーバは、特定の Egress Path に紐づいた EPE 専用アドレスではなく、BGP によってクライアントへの到達性が維持されるデフォルトアドレスで初期コネクションを確立する必要がある。

次に、サーバは各 EPE 専用アドレスをクライアントに通知し、これらのアドレス宛の新たな TCP コネクションが確立される。これらは Subflow として MPTCP セッションに追加される。しかし、前述したようにそれぞれの EPE 専用アドレスは特定の Egress Peer と紐づいたアドレスであるため、実際にコネクションが確立可能で、Subflow として利用できるかどうかは、対応する Egress Peer からクライアント宛の経路を受信しているかに依存する。

4.2.3 MPTCP スケジューラによるトラフィック制御

サーバ上のアプリケーションから送出されたコンテンツトラフィックは、カーネル内で動く MPTCP スケジューラによって各 Subflow に分配される。スケジューラは各 Subflow の TCP メトリクスを基に、その時点で最適な Subflow を選択し、パケットごとに送信先を決定する。すなわち、第 3.2 節で述べた Quality-Based Egress Path Decision Making and Traffic Control のうち、「QoS メトリクス計測」はカーネルの TCP スタックが、「品質の良い EPE Route の選択」と「利用する EPE Route の設定」は MPTCP スケジューラが担う。

パケットを送信する Subflow の選択は MPTCP スケジューラごとの戦略に基づいて行われる。著名なスケジューリング戦略には、“Round-Robin”、“Lowest-RTT-



図 5 Multipath Transport SRv6-EPE Suite でのトラフィックの流れ

Fig. 5 Traffic flow in the Multipath Transport SRv6-EPE Suite.

^{*2} 以後、Egress Peer α から受信した経路は Egress Path α と表現する。

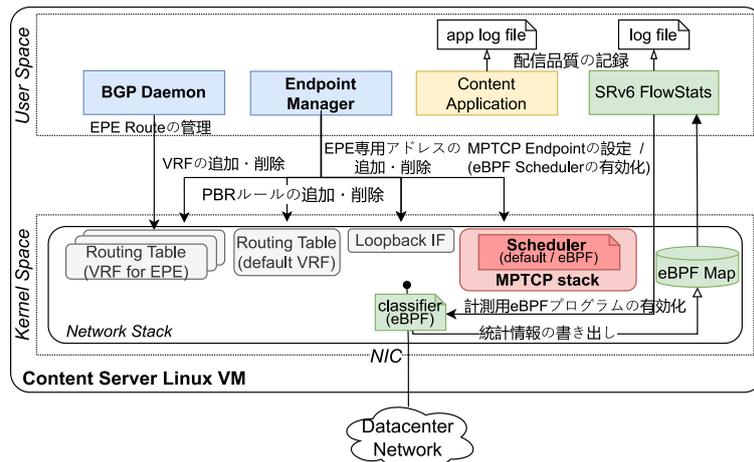


図 6 Multipath Transport SRv6-EPE Suite を実装したコンテンツサーバのコンポーネント群
 Fig. 6 Components of a content server implementing the Multipath Transport SRv6-EPE Suite.

First” [27], および “BLEST” [28] などがある. 本提案手法は, MPTCP スケジューラの戦略をコンテンツの品質基準に合わせて変更することで, 幅広いアプリケーションに最適化する柔軟性を提供する. よって, CSP が本提案手法を展開する際に, コンテンツの種類に合わせて特別に実装および保守しなければならないコンポーネントは MPTCP スケジューラのみになる. これは第 2.4 節で述べた EPE コントロールプレーンの要求事項の 1 つである「展開の容易性」に合致する.

なお, デフォルトアドレスを利用した Subflow (初期コネクション) は, いずれかの EPE 専用アドレスを利用した Subflow と同じ Egress Peer を介して転送される. EPE 専用アドレスを利用した Subflow には SRv6 による処理が追加で発生するため, 仮に BGP ベストパスと一致する EPE Route がパフォーマンスの観点で最適である場合は, 初期コネクションが選択されやすくなることが想定される. 本手法はコンテンツ配信品質の向上を目的としているため, BGP ベストパスを常に利用する Subflow を「品質の良い EPE Route の選択」から排除せず, EPE 専用アドレスによる Subflow と並列して利用するように設計する.

4.3 Multipath Transport SRv6-EPE Suite の実装

第 5 章で後述する評価実験を行うため, Multipath Transport SRv6-EPE Suite に必要なシステムを備えたコンテンツサーバを実装した. 本節では本実装における各コンポーネントの概要を述べる. 図 6 はコンテンツサーバ上で動作するコンポーネント群と, OS 内の関連する機構との相互関係を表したものである.

• MPTCP 実装

第 4.2 節で示したように, 本提案手法では “Quality-Based Egress Path Decision Making and Traffic Control” に MPTCP を活用する. 第 5 章で後述する評価実験では, MPTCPv1 [21] を実装した Linux 仮想

マシンをコンテンツサーバとして利用する. Linux カーネルには 2023 年 10 月時点で最新の開発版のもの (6.6.0-rc4^{*3}) を採用する. なおこの開発版のカーネルでは, TCP の輻輳制御機構をカスタマイズするための eBPF プログラム (BPF STRUCT_OPS) [29] を導入して MPTCP スケジューラを変更することが可能であるが [30], 評価実験にはカーネル内に標準で実装されている BLEST [28] をベースとしたスケジューラ [31] を使用する.

• BGP デーモン

SRv6 L3VPN による “EPE Provisioning and Reachability Management” を実現するためには, CSP AS 内で有効な EPE Route をルートリフレクタから BGP 経由で収集し, サーバ内の各 VRF にインストールする必要がある. 本研究では BGP デーモンに FRRouting^{*4} を利用する.

• Endpoint Manager

第 4.1 節で述べたように, Multipath Transport SRv6-EPE Suite では EPE Route を明示的に利用するために, EPE 専用アドレスとそれぞれのアドレスを VRF にリダイレクトする PBR ルールが必要である. 本研究では EPE 専用アドレスの設定や VRF の作成, MPTCP Endpoint の管理および PBR ルールの適用を一括して行う SDN エージェントを独自に実装した.

5. 評価

本章では, 本稿が提案する Multipath Transport SRv6-EPE Suite が, 第 2.4 節で述べた EPE コントロールプレーンの要求事項を満たしているかどうかを検証する. EPE コントロールプレーンの要求事項には, (1) コンテ

^{*3} 開発コミュニティのリポジトリ (https://github.com/multipath-tcp/mptcp_net-next) で公開されている.

^{*4} FRRouting 9.0.1 を利用. <https://frrouting.org/>

コンテンツ配信品質の向上, (2) 通信環境の変化への追従性, (3) 多様なコンテンツへの柔軟性, (4) 展開の容易性の4つの項目が含まれる。

第4.2節で示したように, 本提案手法は, MPTCPスケジューラの戦略を変更することでEgress Pathの選択基準を変えることが可能である。さらに, 同一ネットワーク上で異なるスケジューラを用いた複数のコンテンツサーバを同時に運用することも可能である。したがって, 品質基準が異なるコンテンツアプリケーションに幅広く適用可能であり, 「(3) 多様なコンテンツへの柔軟性」を有していると言える。

また, 本提案手法は, CSP ASとEgress Peer間で標準的なBGPによる経路交換以外の情報交換を必要としないため, CSPは他のASに協力を仰ぐことなく導入が可能である。加えて第2.2節で述べたように, SRv6-EPEは既存のIPv6ネットワークへ容易に導入でき, EPEを利用しないトラフィックとの共存が可能である。これらの理由から, 本提案手法は「(4) 展開の容易性」の要求を満たすアプローチであると言える。

以後, 本章では本提案手法が「(1) コンテンツ配信品質の向上」と「(2) 通信環境の変化への追従性」を満たすことを定量的に検証するため, 以下のように仮想環境での評価実験を行う。

(1) 実験シナリオ1: ネットワーク品質がサービス提供中に変化しない場合

異なる通信品質を持つEgress Pathが存在する場合において, 本提案手法が適切なEgress Pathを動的に発見および利用することを確認し, 「(1) コンテンツ配信品質の向上」を実現可能か検証する。

(2) 実験シナリオ2: ネットワーク品質がサービス提供中に変化する場合

コンテンツサービス提供中にEgress Pathの通信品質が変化する状況を想定する。本提案手法がこのような環境下でも配信品質の劣化やサービス中断を回避する, 「(2) 通信環境の変化への追従性」を有しているか検証する。

(3) 実験シナリオ3: BGP経路が変化する場合

CSP ASが利用できるEgress Pathに変化が生じた場合でも, EPEコントロールプレーンの2つの役割のうち「EPE Provisioning and Reachability Management」機能が本提案手法に適切に備わっており, EPE Routeが動的に更新されるかを確認する。この実験シナリオでは本提案手法の「(2) 通信環境の変化への追従性」を検証する。

また, 実際のインターネット環境での本提案手法のフィジビリティ評価を目的とした評価実験も行う。WIDEプロジェクト^{*5}が運用するASから, 様々なISPに所属する実験用クライアントに対してコンテンツ配信を行い, より

実環境に近い状態での本提案手法の有用性を実践的に評価する。

5.1 仮想環境での Multipath Transport SRv6-EPE Suite の基本機能の検証

5.1.1 実験環境

図7に示される本評価実験の仮想環境トポロジーは, 第2.1節で設定したトポロジーを単純化し, SRv6 L3VPN経路を集約・配布するルトリフレクタを加えたものである。CSP ASとTransit ISP A/B間, Client ISPとTransit ISP A/B間でEBGPピアを確立し, 経路情報を交換する。CSP ASからClient ISPへのBGPベストパスはTransit A経由とする。以後のシナリオでは, リンクaとbの帯域幅を制限し, 擬似的に各Egress Path間で通信品質の差異が生じている状態を再現する。なお, この仮想環境はLinuxのNetwork Namespace機能[32]によって1台のLinuxマシン上に構築する。各仮想ノードは第4.3節に述べたものと同じBGPデーモンを利用する。

5.1.2 評価用コンテンツトラフィック

本評価実験では, 下記の2つの異なる様式のトラフィックをコンテンツサーバからクライアントに送出させ, それぞれ通常のTCPとSRv6-EPEによるMPTCP(以後“MPTCP EPE”と呼ぶ)を利用した際のスループットを比較分析するために使用する。

- **Constant Bit Rate (CBR) トラフィック**
ライブストリーミングビデオのような一定速度でのコンテンツ配信を想定し, 25 Mbpsのトラフィックをコンテンツサーバからクライアントに向け30秒間送出する。CBRトラフィックの生成にはiperf^{*6}を利用する。
- **Bandwidth-Intensive (BI) トラフィック**
可能な限り帯域を使用するWWWのようなコンテンツを想定し, クライアントはHTTP GETリクエスト

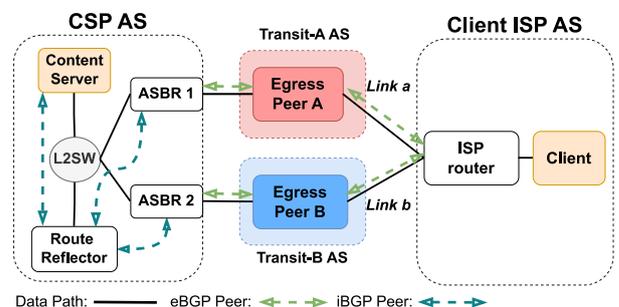


図7 仮想環境を利用した評価実験用トポロジー

Fig. 7 Network topology for the evaluation experiments in a virtual environment.

*5 1988年に設立されたインターネット技術・標準・社会に関連する研究を行うコンソーシアム。

*6 iperf 3.12を利用。https://iperf.fr/

によりコンテンツサーバから 100 MB のバイナリデータをダウンロードする。BI トラフィックの生成には、Python で動作する標準的な WEB サーバ^{*7}を利用する。

5.1.3 SRv6 FlowStats

SRv6-EPE 環境ではコンテンツサーバのアウトバウンドトラフィックは SRv6 でカプセル化されるため、トラフィックが実際にどの EPE Route を使用したかを追跡することが困難である。本評価実験中の MPTCP スケジューラの振る舞いを正確に把握するため、独自のモニタリング用アプリケーション (“SRv6 Flow Stats”) を開発した。

図 6 に示すように、SRv6 Flow Stats はコンテンツサーバ上の Multipath Transport SRv6-EPE Suite の主たるコンポーネントとは独立して動作する。eBPF tc classifier [33] を利用して、コンテンツサーバの物理インタフェースに eBPF プログラムをアタッチし、パケットが通過するごとにインナーパケットの 5-tuple^{*8} をキーとして eBPF Map にカウンタを記録する。ユーザスペースのアプリケーションは 100 ms ごとにこの eBPF Map から統計情報を取得し、差分をログファイルに保存する。これにより、Subflow (TCP コネクション) ごとにトラフィック流量を正確に計測することができる。

5.1.4 実験シナリオ 1：ネットワーク品質がサービス提供中に変化しない場合

本実験では、ネットワークの品質がサービス提供中に変化しない場合での Multipath Transport SRv6-EPE Suite の基本性能を確認する。図 7 のトポロジでのリンク a と b の帯域を制限して、Egress Path の通信品質を 2 種 (高品質パス、低品質パス) 組み合わせた 4 つのパターンを表 1 のように設定した。なお、本節で設定した CBR トラフィックが要求する帯域 (25 Mbps) に対して、高品質パ

表 1 実験シナリオ 1：各パターンのパラメータ

Table 1 Parameters for each pattern in scenario 1.

	リンク a 帯域	リンク b 帯域	BGP ベスト
パターン 1.1: BGP ベストパスが高品質	30 Mbps	15 Mbps	Transit A
パターン 1.2: 両パスともに高品質	30 Mbps	30 Mbps	Transit A
パターン 1.3: BGP ベストパスが低品質	15 Mbps	30 Mbps	Transit A
パターン 1.4: 両パスともに低品質	15 Mbps	15 Mbps	Transit A

^{*7} Python 3.10.12 および標準 HTTP モジュール (<https://docs.python.org/3/library/http.html>) を利用。

^{*8} ここでは、送信元 IP アドレス、送信元ポート番号、宛先 IP アドレス、宛先ポート番号、プロトコル番号を指す。

スではこの要求を満たすリンク帯域 (30 Mbps) を、低品質パスではこの要求に不足するリンク帯域 (15 Mbps) を設定した。

図 8 に、これらのパターンで発生させたコンテンツトラフィックの L7 goodput^{*9} を、TCP と本提案手法における MPTCP (MPTCP EPE) で比較して示す。緑色の点線は CBR の理論上の最大スループットである 25 Mbps を示している。CBR と BI の両方で、MPTCP EPE は通常の TCP と同等以上の L7 goodput を達成している。特にパターン 1.3 や 1.4 のような BGP ベストパスが低品質な状況でも、MPTCP EPE は理論値に近い L7 goodput を記録している。これは Multipath Transport SRv6-EPE Suite が利用可能な Egress Path を効果的に発見し、BGP ベストパスに依存せずに帯域を有効に利用できることを示している。

5.1.5 実験シナリオ 2：ネットワーク品質がサービス提供中に変化する場合

本実験シナリオでは、サービス提供中の配信経路の品質変化に対する Multipath Transport SRv6-EPE Suite の効果を検証する。実験シナリオ 1 と同じネットワークを利用し、テストトラフィックが CBR の場合はサービス開始 4 秒後、BI の場合は 9 秒後に表 2 に示したような品質変化を発生させ、トラフィックの振る舞いを観察した。

図 9 は、サービス開始からの経過時間 (X 軸) ごとの

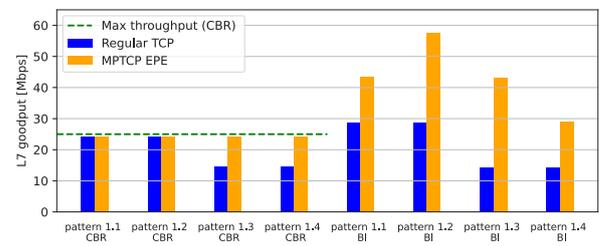


図 8 実験シナリオ 1：各パターンにおける TCP と MPTCP の L7 goodput の比較

Fig. 8 Scenario 1: Comparison of L7 goodput between TCP and MPTCP across different patterns.

表 2 実験シナリオ 2：各パターンのパラメータとその変化

Table 2 Parameters and their changes for each pattern in scenario 2.

	リンク a 帯域	リンク b 帯域	BGP ベスト
パターン 2.1: BGP ベストパスの品質が悪化	30 → 15 Mbps	15 Mbps	Transit A
パターン 2.2: BGP ベストパスの品質が良化	15 → 30 Mbps	15 Mbps	Transit A
パターン 2.3: 冗長パスの品質が悪化	15 Mbps	30 → 15 Mbps	Transit A
パターン 2.4: 冗長パスの品質が良化	15 Mbps	15 → 30 Mbps	Transit A

^{*9} アプリケーションから見たスループット。

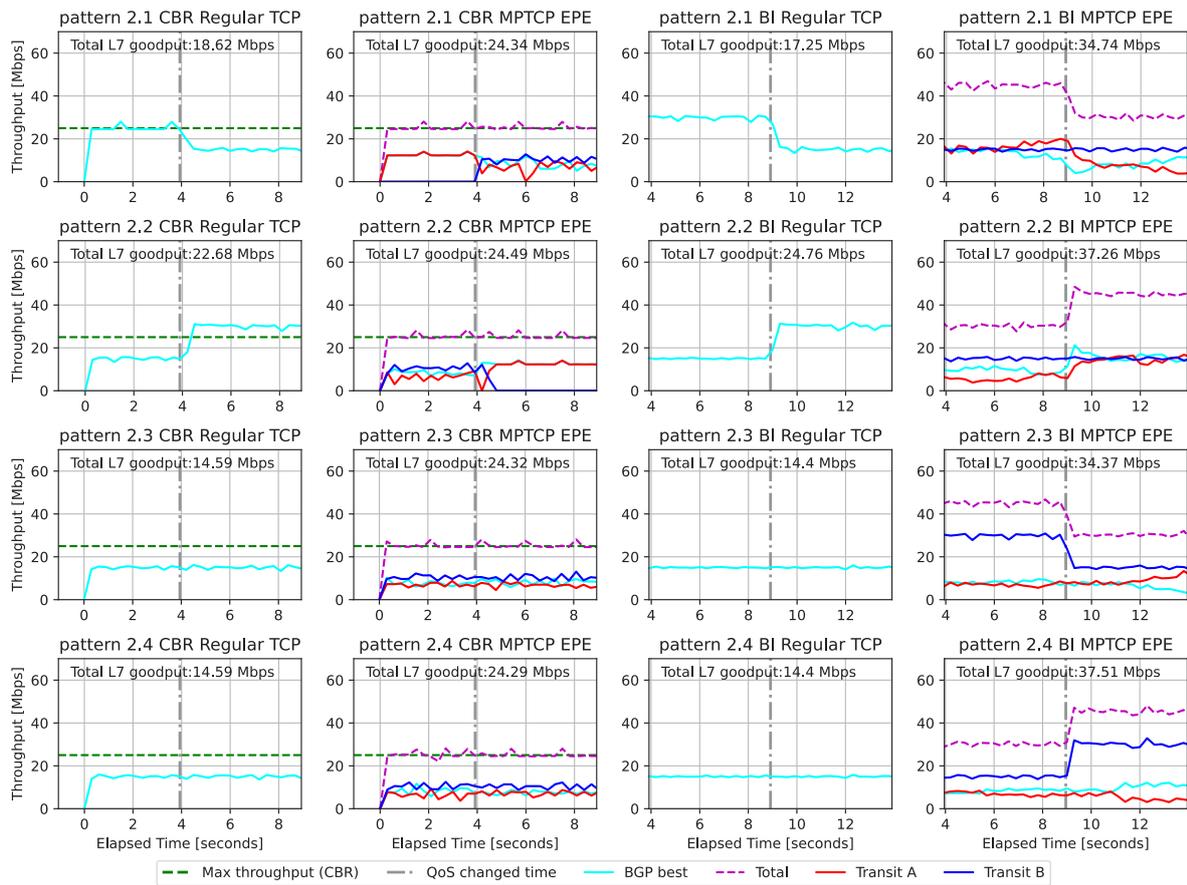


図9 実験シナリオ2：各パターンにおける Egress Path ごとのスループットの推移

Fig. 9 Scenario 2: Temporal throughput changes for each egress path.

各フローのスループット (Y 軸) とセッション全体のスループットの平均値 (各グラフ上部) を、トラフィック種別ごとに表したものである。通常の TCP は BGP ベストパスのみを利用するが、MPTCP EPE では、BGP ベストパス・Transit A・Transit B の3つに対応する Subflow^{*10} が用いられる。赤・青・水色の各線はそれぞれの Subflow のスループットを、紫の線はそれらの合計値を示している。

パターン 2.1 (図9の1行目) では、BGP ベストパスの品質が悪化 (リンク a の帯域減少) した際、通常の TCP は CBR と BI の双方でスループットを落としている。しかし、MPTCP EPE を使用した CBR では、品質変化前は BGP ベストパスと Transit A の Subflow を利用し、変化後は Transit A の代わりに Transit B の Subflow へ切り替え、要求帯域 (25 Mbps) のスループットが維持されている。BI でも、MPTCP EPE は品質変化前後で理論上の最大スループットに近いパフォーマンスを維持し (変化前約

45 Mbps, 変化後約 30 Mbps)、環境変化に適応している。

パターン 2.2 (図9の2行目、BGP ベストパスの品質が良化) では、MPTCP EPE は CBR で要求帯域 (25 Mbps) に近いスループットを維持し、BI では適応的にスループットを増加させている。一方、通常の TCP による CBR は、BGP ベストパスの品質良化前には CBR の要求値に達せず、良化後は要求値を超えるスループットが観測されている。これは品質変化前に TCP によりバッファリングされたデータが、品質変化後に送信された影響と考えられる。また、パターン 2.3 と 2.4 (図9の3、4行目) では、MPTCP EPE は冗長パスを有効に活用し、品質変化後には Transit A・Transit B の双方の合計帯域 (45 Mbps) に近い値のスループットを出していることが分かる。

以上により、Multipath Transport SRv6-EPE Suite は経路上の品質変化に線形に追従し、サービス中断や品質劣化を可能な限り回避する能力を有することが分かった。

5.1.6 実験シナリオ3：BGP 経路が変化する場合

第3.1節で述べたように、適切に EPE を運用するためには、CSP AS 内で利用可能な Egress Path の変化に応じて、コンテンツサーバ上の EPE Route を迅速に更新することが求められる。本実験シナリオでは CSP AS が

^{*10} 第4.2節で述べたように Egress Path が2系統存在する環境下では、(1) BGP ベストパスを利用するデフォルトアドレスによる Subflow, (2) Transit A 由来の Egress Path のみを利用する Subflow, (3) Transit A 由来の Egress Path のみを利用する Subflow の最大3つを利用する。

Egress Peer から受信する BGP 経路が変化する場合を想定し、Multipath Transport SRv6-EPE Suite における SRv6 L3VPN の通信環境変化への追従性を検証する。

図 7 のトポロジにおいて、クライアント ISP のルーターから Transit A に対する経路広告ポリシーが変更（以後経路変更と呼称する）されるケースを仮定する。Egress Peer A がクライアント ISP から受信する経路が変わると、Egress Peer A から CSP の ASBR 1 へ広告される経路にも変更が及ぶ。これに伴って ASBR 1 は自身のルーティングテーブルを更新し、ルータリフレクターを介してコンテンツサーバに経路変更が伝搬され、最終的にコンテンツサーバ上の対応する EPE Route が更新される。

第 4.1 節で述べたように、コンテンツサーバの EPE 専用アドレスによる通信は、該当する Egress Peer に由来する EPE Route を参照して、明示的にその Egress Peer を経由して配送される。コンテンツサーバの EPE 専用アドレスによる Subflow は、MPTCP セッション確立時点でクライアントに到達可能な場合のみ確立される。たとえばコンテンツサービス提供中にクライアントに対する EPE Route が削除された場合、それを参照している EPE 専用アドレスによる Subflow は利用できなくなる。ネットワーク品質がサービス提供中に変化する場合と同様に、MPTCP スケジューラによる他の Subflow へトラフィックの再割り当てが行われる。

本シナリオではクライアント ISP での経路広告ポリシーの変更から、コンテンツサーバが保持する EPE Route へ変更が適用されるまでに掛かる時間を計測し、本提案手法の追従性を定量的に評価する。EPE Route の変更の計測には、ルーティングテーブルの変更をリアルタイムに検知可能な iproute2 の “ip-monitor” 機能 [34] を利用した。また、この経路広告ポリシー変更から、CSP AS 内で BGP ベストパスが収束*11 するまでの時間も合わせて計測し、これらの比較を行う。

本実験シナリオで行った 2 つの経路変更のパターンを表 3 に示す。パターン 3.1 はクライアント ISP が Transit AS A に対して経路を広告していない状態から広告する状態への遷移を、パターン 3.2 は広告している状態から広告しない状態への遷移を想定したものである。それぞれ 100 回ずつ試行し、EPE Route の変更適用と BGP 経路収束の双方を同時に計測した。

表 3 実験シナリオ 3：経路変更のパターン
Table 3 Patterns of route changes in scenario 3.

	BGP ベストパスの変化
パターン 3.1：EPE Route の追加	Transit B → Transit A
パターン 3.2：EPE Route の削除	Transit A → Transit B

*11 CSP AS 内のノードがすべて同一の L2SW で接続されている本トポロジにおいては、コンテンツサーバのルーティングテーブル上で BGP ベストパスの変更が完了された状態と定義する。

東の双方を同時に計測した。

図 10 は、本実験シナリオの各パターンにおいて、コンテンツサーバ上の EPE Route に変更が適用されるまでに掛かった時間と BGP ベストパス収束までの時間の計測結果の分布を、それぞれの四分位数に基づき比較したものである。箱の底辺は第一四分位数を、上辺は第三四分位数を示し、箱の中央線は中央値を表す。ヒゲは外れ値を除いた最小値と最大値の範囲を示し、白い丸点は外れ値を表す。クライアント ISP による経路広告ポリシー変更後、パターン 3.1 における所要時間の中央値はそれぞれ 190 ms 程度、パターン 3.2 においては 310 ms 程度であった。両パターンともに、EPE Route の変更適用と BGP ベストパス収束との間に大きな差異はみられない。

これらの結果より、Multipath Transport SRv6-EPE Suite は通常の IPv6 ルーティングと同等の経路変化への変更追従性を有していることが明らかになった。

5.2 実際のインターネットにおけるコンテンツ配信試験

本節では、Multipath Transport SRv6-EPE Suite のインターネット環境での活用を想定した評価実験に関して述べる。実際のコンテンツ配信に近い環境における本提案手法の有用性と課題を明らかにする。

5.2.1 実験環境

図 11 に本評価実験に利用したネットワークを示す。WIDE プロジェクトが運営する実験用 AS (WIDE AS) にコンテンツサーバを設置する。WIDE AS では 1 台の

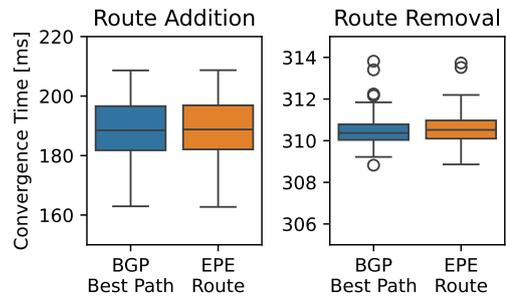


図 10 実験シナリオ 3：BGP 経路の変更の所要時間の分布
Fig. 10 Scenario 3: Distribution of time required for BGP route changes.

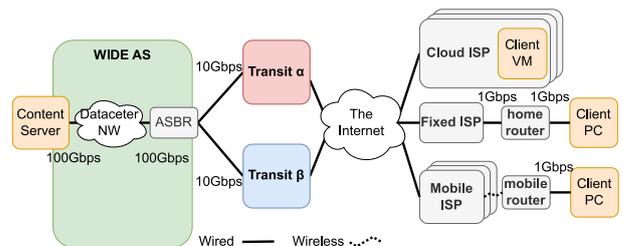


図 11 実際のインターネットの環境を利用した評価実験用トポロジ

Fig. 11 Network topology for the evaluation experiments using the actual internet environment.

ASBR が 2 つの Egress Peer (Transit α , Transit β) と接続している。本実験では様々な ISP に接続したサンプルクライアントを用意し、WIDE AS からコンテンツサービスの提供を行う。WIDE AS 内のコンテンツサーバと ASBR は 100 Gbps で接続しており、本評価実験のために十分な収容能力を持つ。

通常の TCP と MPTCP EPE を利用したコンテンツ配信を同日同時刻帯にそれぞれ 10 回試行し、配信方式による提供品質の差異を比較する。後述する TCP の並列化によるオーバーヘッドを検証するために、本実験では iperf によって生成した BI トラフィックを利用した。通常の TCP では BGP ベストパス選択ルールにより選択された Egress Path を、MPTCP EPE では Transit α と Transit β の両方の Egress Path を経由して配信される。

表 4 に本研究においてサンプルクライアントが接続する各 ISP の基本情報を示す。日本国内の MNO 事業者^{*12} 社と、国内で固定回線でのインターネットアクセスを提供する事業者 1 社、そしてクラウドサービスを提供する国内外の事業者 4 社をクライアントの ISP とした。

表 5 にクライアントデバイスとして利用したホストの詳細なスペックを示す。モバイル事業者および固定回線事業者に接続する際には共通の物理マシンを、クラウド事業者の場合はその事業者が提供している Virtual Private Server (VPS) サービスによる仮想マシンを利用した。本評価実験は ISP 事業者間での性能を直接比較することなく、同一 ISP の配信方式間での提供品質の違いを把握することを目的としているため、実験に利用した仮想マシンのスペックは一致していない。なおすべてのホストの

OS には Ubuntu 22.04.03 LTS を採用した。

5.2.2 通常の TCP と MPTCP EPE 間の L7 goodput の比較

図 12 および図 13 は、各 ISP ごとの配信結果 (L7

表 4 実際のインターネットでのコンテンツ配信試験における各クライアント ISP の概要

Table 4 Overview of client ISPs in the content delivery evaluation on the actual internet.

	BGP ベスト	Transit α AS_PATH 長	Transit β AS_PATH 長	地域と種別
モバイル ISP A	Transit β	3	2	国内 5G
モバイル ISP B	Transit α	2	3	国内 5G
モバイル ISP C	Transit α	2	3	国内 5G
モバイル ISP D	Transit α	2	2	国内 5G
固定回線 ISP E	Transit α	2	2	国内 FTTH
クラウド ISP F	Transit β	5	3	米国 クラウド
クラウド ISP G	Transit α	2	2	国内 クラウド
クラウド ISP H	Transit β	3	2	国内 クラウド
クラウド ISP I	Transit β	5	4	国内 クラウド

表 5 実際のインターネットでのコンテンツ配信試験に利用したコンテンツサーバとクライアントデバイスのスペック

Table 5 Specifications of content servers and client devices used in the content delivery evaluation on the actual internet.

種別	CPU	コア数	メモリ	Kernel ver.	
コンテンツサーバ	仮想	AMD EPYC 7713	16	16GB	6.6.0-rc4
モバイル ISP A	物理	AMD Ryzen 5 PRO 2400GE	4	30GB	6.2.0-36-generic
モバイル ISP B	物理	AMD Ryzen 5 PRO 2400GE	4	30GB	6.2.0-36-generic
モバイル ISP C	物理	AMD Ryzen 5 PRO 2400GE	4	30GB	6.2.0-36-generic
モバイル ISP D	物理	AMD Ryzen 5 PRO 2400GE	4	30GB	6.2.0-36-generic
固定回線 ISP E	仮想	Intel Core Processor (Broadwell)	1	1GB	5.15.0-84-generic
クラウド ISP F	仮想	Intel Xeon E5630	2	4GB	5.15.0-79-generic
クラウド ISP G	仮想	Intel(R) Xeon(R) Gold 6121U CPU	4	4GB	5.15.0-53-generic
クラウド ISP H	仮想	Intel Xeon Processor (Cascadelake)	1	1GB	5.15.0-84-generic
クラウド ISP I	仮想	Intel Xeon Processor (Cascadelake)	1	1GB	5.15.0-84-generic

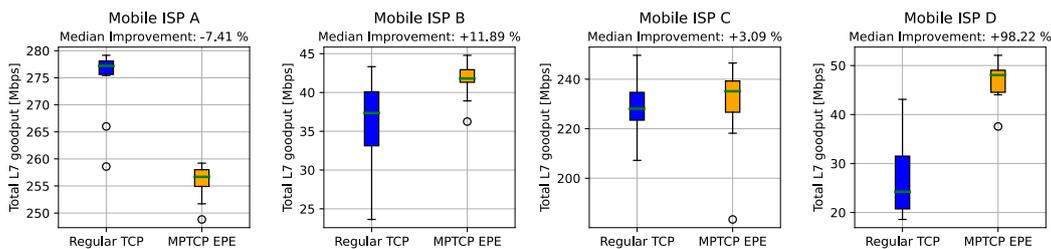


図 12 各モバイル ISP の L7 goodput
Fig. 12 L7 goodput of each mobile ISP.

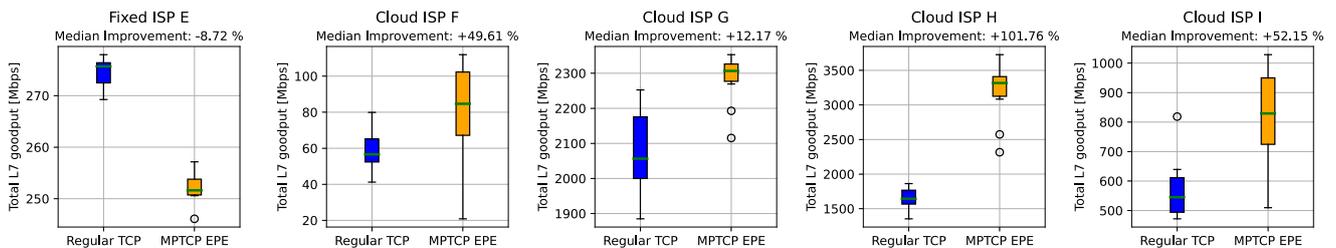


図 13 各固定回線・クラウド ISP の L7 goodput
Fig. 13 L7 goodput of each fixed line and cloud ISP.

*12 移動通信サービスに係る無線局を自ら運用している事業者

goodput) の分布を計測種別 (通常の TCP もしくは MPTCP EPE) ごとに四分位数に基づき表したものである。箱の底辺は第一四分位数を、上辺は第三四分位数を示し、緑の線は中央値を表す。ヒゲは外れ値を除いた最小値と最大値の範囲を示し、白い丸点は外れ値を表している。また、通常の TCP に対する MPTCP EPE の改善割合 (中央値比) を各グラフ上部に示している。

全体を通して、本実験環境では多くの ISP で、MPTCP EPE のほうがより高品質なサービス提供を行えていることがみて取れる。9つの ISP 中6つで10%以上の品質向上をみせたほか、内4つのISPでは50%以上の顕著な向上がみられた。

しかしながら ISP A および ISP E では、MPTCP EPE を使用することにより通常の TCP より約8%ほど品質が悪化している。これはEPEでより優れたEgress Pathを選択することで得られる利得よりも、MPTCP EPEに掛かる処理のオーバーヘッド、あるいはTCPフローの並列化による負荷が大きくなっていることが推察される。

5.2.3 ISP EにおけるTCPコネクション並列化の負荷検証

品質悪化率が最も高かったISP Eを対象とし、通常のTCPを並列化して同様の条件でコンテンツを提供する追加実験を行った。本実験ではiperfコマンドの並列化オプション^{*13}を利用して、通常のTCPによるコンテンツ配信のコネクション並列化を行う。

図14に、ISP Eに接続したクライアントに対して、iperfのデータ用TCPコネクションを1, 3, 6, 9, 12に並列化してコンテンツ配信を行った際のそれぞれのL7 goodputを示す。最も左側に、参考までにMPTCP EPEでの結果を追記している。TCPコネクションを並列化す

るに従って、L7 goodputが逡減していることが確認できる。

今回の実験環境で利用したiperfでは、実際の計測用トラフィックが乗るTCPコネクションの他に、メトリクスを相互に交換するための制御用のTCPコネクションを確立する[35]。本実験環境におけるMPTCP EPEは、1つのMPTCPセッションで合計3つのSubflow (BGPベストパスを利用するSubflow, Transit α のみを利用するSubflow, Transit β のみを利用するSubflow) が確立されるため、MPTCPを利用するには6つのTCPコネクションが確立されることになる。

図14より、合計6つのTCPコネクションを確立するMPTCP EPEと、合計7つ (制御用TCPコネクションと6つのデータ用TCPコネクション) のTCPコネクションを利用した際のL7 goodputはおおよそ近い値になっていることが確認できる。すなわち、ISP Eの環境下ではTCPコネクションを並列化することによる負荷が顕著であり、これに伴ってMPTCP EPEのスループットが低下していると言える。

本追加実験を通じて、MPTCP EPEを利用してユーザーにコンテンツを提供することが、必ずしも提供品質の改善につながるわけではないことが明らかになった。

5.2.4 まとめ

本評価実験を通じて、Multipath Transport SRv6-EPE Suiteが実際のインターネットでのコンテンツ配信環境においても、多くの場合で配信品質向上を見込めることが明らかになった。一方で、一部の環境下においては、TCPコネクションが並列化されることによって配信品質の低下を招く可能性があることが分かった。

6. 関連研究

本章では、本提案手法と提案手法の要素技術 (EPE と Multipath Transport) に関連する研究とを比較し、提案手法の特徴をより明らかにする。

6.1 Egress Peer Engineering

EPEは、BGPベストパスセレクションによらずに、AS外に向かうトラフィックを柔軟に制御する技術である。一般的な他のトラフィックエンジニアリング技術と比較して、自らが管理するネットワーク外の通信経路の選択にフォーカスしている点が特徴的である。

6.1.1 EPEの活用

EPEは、その特徴から様々な用途に用いることができる。Espresso [36]では対外接続リンクの輻輳回避を、BGP-NCM [37]では帯域の有効活用を目的とし、EPEが用いられている。これらの手法は他のASとの接続リンクの状態を基準にEgress Pathを選択する取り組みであり、我々の手法のようにクライアントまでのE2Eの品質向上

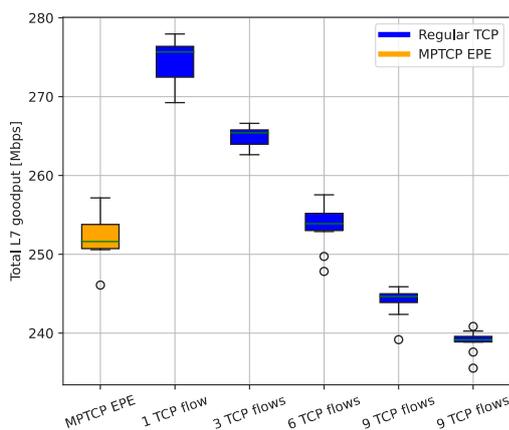


図14 ISP Eにおける並列化したTCPとMPTCP EPEのL7 goodputの比較

Fig. 14 Comparison of L7 goodput between parallelized TCP and MPTCP EPE in ISP E.

^{*13} iperf3 -P 並列数のように実行することでデータ用TCPコネクションを並列化できる。

を目的としたものではない。

また、中村らの研究 [18] や筆者らの研究 [12] は、インターネットの品質測定を目的としており、EPEを複数のBGPパスを比較するために利用している。これらの手法によってE2Eでの性能測定をもとにEgress Pathの評価ができる一方、実際にパフォーマンスに優れると判断されたEgress Pathの具体的な活用手法は取り組みのスコープとされていない。

EPEの具体的な制御手法として、BGP Link State (BGP-LS) のプロトコル拡張 [38] と、それをを用いた運用モデルが標準化されている [11]。しかしながらこれらの標準はあくまでEPEの集中管理を行う際の一般論についてのみ記述しており、具体的な配信品質の向上手法や、経路や品質の変更に追従する手法に関する検討が不足している。

6.1.2 EPEを用いたコンテンツの品質向上

EPEでは、トラフィックの配信に利用するEgress Pathを任意に切り替えることで、クライアントまでのE2Eの通信品質を変化させることができる。この特性を利用し、通信品質の向上を目指す取り組みがいくつか存在する。Edge Fabric [39] は、外部接続リンクの有効活用とコスト最小化を主目的としているものの、実際のプロダクションコンテンツの配信品質をベースにEgress Pathに重み付けを行う。筆者らの過去の研究であるCOFFEE [40] では、データプレーンにSRv6を利用し、Kubernetesベースのコントローラを利用することでEPEによるコンテンツ配信品質の向上を実現している。

しかしながらこれらの手法にも未解決の課題がある。Facebookの手法はAS全体で統一して経路選択戦略を決定するため、本稿での提案手法のように、コンテンツやクライアントに合わせて選択基準を柔軟に組み替えることができない。COFFEEは経路制御のために専用のSDNフレームワークが必要となるため、開発・導入・運用に掛かるコストが高く、第2.4節で定義したEPEコントロールプレーンの要求事項である「展開の容易性」に課題がある。本稿で提案するMultipath Transport SRv6-EPE Suiteは主要なコンポーネントに既存のプロトコル実装を活用しているため、COFFEEと比較して開発・導入・運用に掛かるコストが低い。また、COFFEEはセッションの開始前に過去の統計情報からトラフィックを送出するEgress Pathを決定するため、サービス中にEgress Pathの通信品質が変動した場合に、サービス断や配信品質の劣化を招く可能性がある。Multipath Transport SRv6-EPE SuiteではリアルタイムにEgress Pathが評価・選択されるため、サービス中の通信品質の変化に追従できる。

6.2 Multipath Transport

Multipath Transportは、トランスポート層でコネク

ションの多重化とスケジューリングを行うプロトコルの総称である [41], [42]。代表的なプロトコルとして、MPTCP [21], MP-QUIC [43], MP-DCCP [44] などが広く利用されている。

6.2.1 Multipath Transport の活用

一般に、Multipath Transportは複数のリンクの帯域のアグリゲーションや冗長化を目的として利用される。Multipath Transportが複数の経路を併用できる性質に着目し、帯域だけでなく、遅延やジッタなど様々なQoSメトリクスを向上させる取り組みが存在する。

たとえばNikravesららの研究などでは [45], MPTCPを用いて複数のモバイル回線を束ねることで、スループットと冗長性の向上を実現している。Hurtigららの研究の例 [46] では、モバイル回線とWi-Fiなど、複数のインタフェースを持つ機器にMPTCPを適用しWebコンテンツを表示するまでの遅延を削減している。また、このようなアプローチは他のプロトコルを利用しても実践されている [47]。

金子らの研究 [48] では、WebRTCにより複数のモバイル回線に映像を重複して配送することで、自動車のロバストな遠隔監視アプリケーションを実装している。

X-LINK [49] は、多くの動画コンテンツによって構成されたeコマースサービスのQoE向上を目的として、MP-QUICを拡張し大規模なデプロイメントを行った研究である。クライアントアプリケーションからのQoEコントロールシグナルによって、コンテンツサーバのスケジューラがコンテンツを送出する宛先アドレスを選択する機構が導入されており、サービス応答性やスタビリティの向上が実現されている。

これらの既存の研究がフォーカスしているのはクライアントデバイスのアクセス回線の冗長化・抽象化であり、Multipath Transport SRv6-EPE Suiteのように品質に応じたBGP経路の選択と利用に用いることを目的としていない。

また、QualitySDN [50] では、SRネットワーク上でMPTCPサブフローを転送することにより、QoEを考慮したE2Eの転送を行っているが、制御が可能な範囲はSRドメインの中に限られており、我々の研究のように、管理ドメイン下でない通信経路上での利用はできない。

6.2.2 Multipath Transport スケジューラ

Multipath Transportには、トラフィックを送出するSubflowを選ぶスケジューラを組み替えることで、通信品質の更なる向上を目指す余地があると言われている [51]。

たとえばLinux Kernelには、Subflowを一定の順序で利用するものや複数のパスに必ず同時送信するもののようなプリミティブなスケジューラから、経路間での公平なデータ配分を目的としたwVegas [52], ヘテロロニアスな品質のアクセス回線でのパフォーマンス向上を目的とした

BLEST [28] や ECF [53] など、多種多様なスケジューラがこれまでに実装されてきた。

これらのアプローチは幅広いアプリケーションに広く利用可能なものであるが、近藤ら [54] が既存の一般的なスケジューラのビデオストリーミングでの問題を提起しているように、一部のアプリケーションでは従来のトランスポート層のメトリクスのみを利用したスケジューラを用いることが限界を来していると指摘されている [55]。そのため近年ではアプリケーションのユースケースに特化させたスケジューラ実装も広く提案されている。

たとえばビデオストリーミングの領域では、アプリケーション層や動画ストリーミングのメトリクスを利用した新しいアプローチが提案されている [55], [56]。また、Virtual Reality 用のユースケースに特化したスケジューリングを行うフレームワークも提案されている [57]。

このように様々なスケジューラがこれまでに提案されているが、MPTCP を利用する Multipath Transport SRv6-EPE Suite でも、アプリケーションのユースケースに合わせたスケジューラ戦略を導入できる。第 4.3 節でも述べたように、本研究で利用した最新の Linux カーネルの開発版では、eBPF を用いて運用者が独自のスケジューラを導入することができる。たとえば、eBPF Map [58] によりユーザランドの任意のアプリケーションからスケジューリングに介入し、個別のコンテンツアプリケーションのメトリクスを利用したスケジューリング戦略を採用するような使い方も想定される。

7. 今後の展望

本研究では Multipath Transport のプロトコルとして、Linux カーネルに標準的に実装されている MPTCP を選定したが、他のプロトコルを用いた際の性能評価は今後の課題である。特に、Multipath QUIC は MPTCP と比較してパケット損失に対して優れた結果を示しているとの研究があり [59]、プロトコルの違いによる振る舞いの違いについての検証が期待される。加えて、従来の Multipath Transport のユースケースのように、クライアントデバイスが複数のアクセス回線を有する場合での本提案手法のパフォーマンスに関しても、さらなる実験と考察が必要となるだろう。

また、実際のコンテンツ配信では、アプリケーションごとに異なる品質基準が要求される場合がある。本研究が提案した Multipath Transport SRv6-EPE Suite では、MPTCP スケジューラを変更することにより、異なる要求に応じた柔軟な経路選択が可能である。ユーザの体感品質に強く関連する QoS メトリクスを明らかにし、そのようなメトリクスを利用した適切な MPTCP スケジューラを選択することが求められる。第 6 章でも述べたように特定のアプリケーションに最適化されたスケジューラに関する

研究は数多く行われているが、これらのスケジューラを Multipath Transport SRv6-EPE Suite にて利用した際の有用性の評価が今後の課題となる。

加えて、本研究での評価実験を通じて、本提案手法を利用することにより逆に配信品質の劣化を招く環境が一部に存在することも明らかになっている。今後はより多様な環境で本提案手法の実証実験を行い、パフォーマンス劣化の原因を明らかにしていく。それに基づき、クライアント ISP の環境の特性を判断してコンテンツ提供方式をダイナミックに切り替える機構を取り入れることで、本提案手法をより幅広く利用することができるようになるだろう。

そして筆者らの過去の研究 [12] では、共通する ISP のユーザ間で、適した Egress Path が共通する傾向があることが分かっている。クライアント ISP ごとに統計情報を蓄積することで、より効率的に Egress Path を選択する技術も考えられる。

8. 結論

本研究では、CSP のネットワークにおいてパフォーマンスに基づいた経路選択を行うための機構を「EPE コントロールプレーン」と定義し、必要な機能を明確にしたうえで、MPTCP と SRv6 L3VPN を利用した “Multipath Transport SRv6-EPE Suite” を提案した。

本提案手法の基本性能を検証するために、CSP のネットワークを模した仮想環境での評価実験を行った。本評価実験では BGP ベストパスのみを利用した従来のコンテンツ配信と、Multipath Transport SRv6-EPE Suite によるコンテンツ配信のパフォーマンスを比較し、すべてのテストケースにおいて本提案手法が従来のコンテンツ配信の同等以上の性能を有することが確認できた。また、いくつかのケースにおいてはスループットが大きく向上し、かつサービスの安定性向上に寄与することが確かめられた。

さらに、日本の主要なモバイル ISP4 社を含む 9 つの ISP に接続するサンプルクライアントに対して、インターネットを介してコンテンツサービスを提供する実環境での評価実験も実施した。本実験環境においては、9 つ中 6 つの ISP に接続したサンプルクライアントで、従来のコンテンツ配信環境と比較してスループットが大きく向上することが明らかとなった。

謝辞 本研究について、日頃から助言を賜りました WIDE プロジェクトや vSIX WG のメンバー、実験リソースを提供いただいたさくらインターネット株式会社、株式会社ブロードバンドタワー、NTT コミュニケーションズ株式会社にこの場をお借りし感謝申し上げます。

参考文献

- [1] Akamai: Measuring Video Quality and Performance: Best Practices, Technical report, Akamai (2020).

- [2] Schlinker, B., Cunha, I., Chiu, Y.-C., Sundaresan, S. and Katz-Bassett, E.: Internet Performance from Facebook's Edge, *Proc. of the Internet Measurement Conference (IMC '19)*, New York, NY, USA, Association for Computing Machinery, pp.179–194 (online), DOI: 10.1145/3355369.3355567 (2019).
- [3] Rekhter, Y., Hares, S. and Li, T.: A Border Gateway Protocol 4 (BGP-4), RFC 4271 (2006).
- [4] Li, Z., Levin, D., Spring, N. and Bhattacharjee, B.: Internet Anycast: Performance, Problems, & Potential, *Proc. of the 2018 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '18)*, New York, NY, USA, Association for Computing Machinery, pp.59–73 (online), DOI: 10.1145/3230543.3230547 (2018).
- [5] Savage, S., Collins, A., Hoffman, E., Snell, J. and Anderson, T.: The End-to-End Effects of Internet Path Selection, *Proc. of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM '99)*, New York, NY, USA, Association for Computing Machinery, pp.289–299 (online), DOI: 10.1145/316188.316233 (1999).
- [6] Xiao, L., Lui, K.-S., Wang, J. and Nahrstedt, K.: QoS extension to BGP, *10th IEEE International Conference on Network Protocols, 2002. Proceedings.*, pp.100–109 (online), DOI: 10.1109/ICNP.2002.1181390 (2002).
- [7] Egilmez, H. E. and Tekalp, A. M.: Distributed QoS Architectures for Multimedia Streaming Over Software Defined Networks, *IEEE Trans. on Multimedia*, Vol.16, No.6, pp.1597–1609 (online), DOI: 10.1109/TMM.2014.2325791 (2014).
- [8] obinata: eSports 時代の ISP を考える : Troubleshooting Gaming network connectivity, <https://www.janog.gr.jp/meeting/janog51/wp-content/uploads/2022/12/janog51-esports-obinata.pdf> (2023).
- [9] Liotou, E., Tsolkas, D. and Passas, N.: A roadmap on QoE metrics and models, *2016 23rd International Conference on Telecommunications (ICT)*, pp.1–5 (online), DOI: 10.1109/ICT.2016.7500363 (2016).
- [10] Tsolkas, D., Liotou, E., Passas, N. and Merakos, L.: A survey on parametric QoE estimation for popular services, *Journal of Network and Computer Applications*, Vol.77, pp.1–17 (online), DOI: <https://doi.org/10.1016/j.jnca.2016.10.016> (2017).
- [11] Filsfils, C., Previdi, S., Dawra, G., Aries, E. and Afanasiev, D.: Segment Routing Centralized BGP Egress Peer Engineering, RFC 9087 (2021).
- [12] Toyota, Y., Mishima, W., Kanaya, K. and Nakamura, O.: Performance Aware Egress Path Discovery for Content Provider with SRv6 Egress Peer Engineering, *IEICE Trans. on Information and Systems*, Vol.E106.D, No.5, pp.927–939 (online), DOI: 10.1587/transinf.2022NTP0003 (2023).
- [13] Filsfils, C., Previdi, S., Ginsberg, L., Decraene, B., Litkowski, S. and Shakir, R.: Segment Routing Architecture, RFC 8402 (2018).
- [14] Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsuhashima, S. and Li, Z.: Segment Routing over IPv6 (SRv6) Network Programming, RFC 8986 (2021).
- [15] Humble, J. and Farley, D.: *Continuous delivery: reliable software releases through build, test, and deployment automation*, Pearson Education (2010).
- [16] Faratin, P., Clark, D. D., Bauer, S. and Lehr, W.: Complexity of Internet interconnections: Technology, incentives and implications for policy (2007).
- [17] Jager, M.: Securing IXP connectivity, *Proc. APNIC*, pp.1–41 (2012).
- [18] Nakamura, R., Shimizu, K., Kamata, T. and Pelsser, C.: A First Measurement with BGP Egress Peer Engineering, *International Conference on Passive and Active Network Measurement*, Springer, pp.199–215 (2022).
- [19] Dawra, G., Talaulikar, K., Raszuk, R., Decraene, B., Zhuang, S. and Rabadan, J.: BGP Overlay Services Based on Segment Routing over IPv6 (SRv6), RFC 9252 (2022).
- [20] Faucheur, F. L., Clercq, J. D., Ooms, D. and Carugi, M.: BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN, RFC 4659 (2006).
- [21] Ford, A., Raiciu, C., Handley, M. J., Bonaventure, O. and Paasch, C.: TCP Extensions for Multipath Operation with Multiple Addresses, RFC 8684 (2020).
- [22] Habib, S., Qadir, J., Ali, A., Habib, D., Li, M. and Sathiseelan, A.: The past, present, and future of transport-layer multipath, *Journal of Network and Computer Applications*, Vol.75, pp.236–258 (online), DOI: <https://doi.org/10.1016/j.jnca.2016.09.005> (2016).
- [23] Raiciu, C., Barre, S., Pluntke, C., Greenhalgh, A., Wischik, D. and Handley, M.: Improving Datacenter Performance and Robustness with Multipath TCP, *Proc. of the ACM SIGCOMM 2011 Conference (SIGCOMM '11)*, New York, NY, USA, Association for Computing Machinery, pp.266–277 (online), DOI: 10.1145/2018436.2018467 (2011).
- [24] Chirwa, R. M. N. and Lauf, A. P.: Performance improvement of transmission in Unmanned Aerial Systems using multipath TCP, *2014 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp.000019–000024 (online), DOI: 10.1109/ISSPIT.2014.7300557 (2014).
- [25] Chao, L., Wu, C., Yoshinaga, T., Bao, W. and Ji, Y.: A Brief Review of Multipath TCP for Vehicular Networks, Sensors, Vol.21, No.8 (online), DOI: 10.3390/s21082793 (2021).
- [26] Łuczak, c. P., Ignaciuk, P. and Morawski, M.: Evaluating MPTCP Congestion Control Algorithms: Implications for Streaming in Open Internet, *Future Internet*, Vol.15, No.10 (online), DOI: 10.3390/fi15100328 (2023).
- [27] Paasch, C., Ferlin, S., Alay, O. and Bonaventure, O.: Experimental Evaluation of Multipath TCP Schedulers, *Proc. of the 2014 ACM SIGCOMM Workshop on Capacity Sharing Workshop (CSWS '14)*, New York, NY, USA, Association for Computing Machinery, pp.27–32 (online), DOI: 10.1145/2630088.2631977 (2014).
- [28] Ferlin, S., Alay, c., Mehani, O. and Boreli, R.: BLEST: Blocking estimation-based MPTCP scheduler for heterogeneous networks, *2016 IFIP Networking Conference (IFIP Networking) and Workshops*, pp.431–439 (online), DOI: 10.1109/IFIPNetworking.2016.7497206 (2016).
- [29] Hinz, J.-T., Addanki, V., Györgyi, C., Jepsen, T. and Schmid, S.: TCP's Third Eye: Leveraging EBPF for Telemetry-Powered Congestion Control, *Proc. of the 1st Workshop on EBPF and Kernel Extensions (eBPF '23)*, New York, NY, USA, Association for Computing Machinery, pp.1–7 (online), DOI: 10.1145/3609021.3609295 (2023).
- [30] Multipath-Tcp: Linux MPTCP Upstream Project, https://github.com/multipath-tcp/mptcp_net-next/wiki. Accessed: 2023-11-19.
- [31] Multipath-Tcp: Linux MPTCP default scheduler code, https://github.com/multipath-tcp/mptcp_net-next/blob/d2edcbba43b5b9b281e71265481a4498443a636a/net/

- mptcp/protocol.c#L1449-L1459). Accessed: 2024-02-03.
- [32] Biederman, E. W. and Networx, L.: Multiple instances of the global linux namespaces, *Proc. of the Linux Symposium*, Vol.1, No.1, Citeseer, pp.101–112 (2006).
- [33] Borkmann, D.: On getting tc classifier fully programmable with cls bpf, *Proc. of netdev*, Vol.1 (2016).
- [34] Debian iproute2 Team: ip-monitor(8) — iproute2 — Debian unstable — Debian Manpages, (<https://manpages.debian.org/unstable/iproute2/ip-monitor.8.en.html>) (2012) (accessed: 2024-02-12).
- [35] ESnet: iperf3 Manual Page, (<http://software.es.net/iperf/invoking.html#iperf3-manual-page>) (accessed: 2023-11-19).
- [36] Yap, K.-K., Motiwala, M., Rahe, J., Padgett, S., Holliman, M., Baldus, G., Hines, M., Kim, T., Narayanan, A., Jain, A. et al.: Taking the edge off with espresso: Scale, reliability and programmability for global internet peering, *Proc. of the Conference of the ACM Special Interest Group on Data Communication*, pp.432–445 (2017).
- [37] Moubarak, M., Elbayoumy, A. D. and Megahed, M. H.: Design and implementation of BGP novel control mechanism (BGP-NCM) based on network performance parameters, *Ain Shams Engineering Journal*, Vol.9, No.4, pp.2079–2091 (2018).
- [38] Previdi, S., Talaulikar, K., Filsfils, C., Patel, K., Ray, S. and Dong, J.: Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing BGP Egress Peer Engineering, RFC 9086 (2021).
- [39] Schlinker, B., Kim, H., Cui, T., Katz-Bassett, E., Madhyastha, H. V., Cunha, I., Quinn, J., Hasan, S., Lapukhov, P. and Zeng, H.: Engineering egress with edge fabric: Steering oceans of content to the world, *Proc. of the Conference of the ACM Special Interest Group on Data Communication*, pp.418–431 (2017).
- [40] Kanaya, K., Toyota, Y., Mishima, W., Shirokura, H. and Esaki, H.: Qoe-aware content oriented path optimization framework with egress peer engineering, *2022 Tenth International Symposium on Computing and Networking (CANDAR)*, IEEE, pp.36–45 (2022).
- [41] Habib, S., Qadir, J., Ali, A., Habib, D., Li, M. and Sathiaselan, A.: The past, present, and future of transport-layer multipath, *Journal of Network and Computer Applications*, Vol.75, pp.236–258 (2016).
- [42] Xu, C., Zhao, J. and Muntean, G.-M.: Congestion control design for multipath transport protocols: A survey, *IEEE communications surveys & tutorials*, Vol.18, No.4, pp.2948–2969 (2016).
- [43] Liu, Y., Ma, Y., Coninck, Q. D., Bonaventure, O., Huitema, C. and Kuhlwind, M.: Multipath Extension for QUIC, Internet-Draft draft-ietf-quic-multipath-06, Internet Engineering Task Force (2023). Work in Progress.
- [44] Amend, M., Brunstrom, A., Kessler, A., Rakocevic, V. and Johnson, S.: DCCP Extensions for Multipath Operation with Multiple Addresses, Internet-Draft draft-ietf-tsvwg-multipath-dccp-11, Internet Engineering Task Force (2023). Work in Progress.
- [45] Nikraves, A., Guo, Y., Qian, F., Mao, Z. M. and Sen, S.: An in-depth understanding of multipath TCP on mobile devices: Measurement and system design, *Proc. of the 22nd Annual International Conference on Mobile Computing and Networking*, pp.189–201 (2016).
- [46] Hurtig, P., Grinnemo, K.-J., Brunstrom, A., Ferlin, S., Alay, Ö. and Kuhn, N.: Low-latency scheduling in MPTCP, *IEEE/ACM Transactions on Networking*, Vol.27, No.1, pp.302–315 (2018).
- [47] Viernickel, T., Froemmgen, A., Rizk, A., Koldehofe, B. and Steinmetz, R.: Multipath QUIC: A deployable multipath transport protocol, *2018 IEEE International Conference on Communications (ICC)*, IEEE, pp.1–7 (2018).
- [48] 金子直矢, 伊東孝紘, 勝田 肇, 渡辺敏暢, 阿部 博, 大西亮吉ほか: 複数回線を冗長併用する通信技術の WebRTC 映像伝送への適用と評価, *情報処理学会論文誌 デジタルプラクティス (TDP)*, Vol.3, No.3, pp.21–31 (2022).
- [49] Zheng, Z., Ma, Y., Liu, Y., Yang, F., Li, Z., Zhang, Y., Zhang, J., Shi, W., Chen, W., Li, D. et al.: Xlink: Qoe-driven multipath quic transport in large-scale video services, *Proc. of the 2021 ACM SIGCOMM 2021 Conference*, pp.418–432 (2021).
- [50] Barakabitze, A. A., Mkwawa, I.-H., Sun, L. and Ifeachor, E.: QualitySDN: Improving video quality using MPTCP and segment routing in SDN/NFV, *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, IEEE, pp.182–186 (2018).
- [51] Yang, F., Amer, P. and Ekiz, N.: A scheduler for multipath TCP, *2013 22nd International Conference on Computer Communication and Networks (ICCCN)*, IEEE, pp.1–7 (2013).
- [52] Cao, Y., Xu, M. and Fu, X.: Delay-based congestion control for multipath TCP, *2012 20th IEEE International Conference on Network Protocols (ICNP)*, pp.1–10 (online), DOI: 10.1109/ICNP.2012.6459978 (2012).
- [53] Lim, Y.-s., Nahum, E. M., Towsley, D. and Gibbens, R. J.: ECF: An MPTCP Path Scheduler to Manage Heterogeneous Paths, *Proc. of the 13th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '17, New York, NY, USA, Association for Computing Machinery, p.147–159 (online), DOI: 10.1145/3143361.3143376 (2017).
- [54] 近藤優吉, カベンディッシュジュルセウ, 野林大起, 池永全志: モバイルネットワークにおけるビデオストリーミングのための MPTCP スケジューラの性能評価, *電子情報通信学会技術研究報告. IA, インターネットアーキテクチャ*, Vol.121, No.68, pp.62–67 (オンライン), 入手先 (<https://cir.nii.ac.jp/crid/1050571859119409024>) (2021).
- [55] Corbillon, X., Aparicio-Pardo, R., Kuhn, N., Texier, G. and Simon, G.: Cross-Layer Scheduler for Video Streaming over MPTCP, *Proceedings of the 7th International Conference on Multimedia Systems*, MMSys '16, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/2910017.2910594 (2016).
- [56] Xing, Y., Xue, K., Zhang, Y., Han, J., Li, J., Liu, J. and Li, R.: A Low-Latency MPTCP Scheduler for Live Video Streaming in Mobile Networks, *IEEE Trans. on Wireless Communications*, Vol.20, No.11, pp.7230–7242 (online), DOI: 10.1109/TWC.2021.3081498 (2021).
- [57] Wei, W., Han, J., Xing, Y., Xue, K., Liu, J. and Zhuang, R.: MP-VR: An MPTCP-Based Adaptive Streaming Framework for 360-degree Virtual Reality Videos, *ICC 2021 - IEEE International Conference on Communications*, pp.1–6 (online), DOI: 10.1109/ICC42927.2021.9500817 (2021).
- [58] kernel development community, T.: Linux kernel docs BPF maps, (<https://docs.kernel.org/bpf/maps.html>). Accessed: 2024-02-03.
- [59] De Coninck, Q. and Bonaventure, O.: Multipath QUIC: Design and Evaluation, *Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '17, New York, NY, USA, Association for Computing Machinery, pp.160–166 (on-

line), DOI: 10.1145/3143361.3143370 (2017).



豊田 安信 (学生会員)

2020年 慶應義塾大学政策・メディア研究科修士課程修了。同年より同大学同研究科博士課程在学中。(株)ブロードバンドタワー所属。インターネット運用技術やSDNなどの研究に従事。WIDEプロジェクトボードメンバー。



三島 航

2019年 北陸先端科学技術大学院大学修了。同年よりNTTコミュニケーションズ(株)に所属。SR技術をはじめとする経路制御、Streaming Telemetry・IPFIX等の監視技術の開発に従事。WIDEプロジェクトメンバー。



早坂 彪流

2021年 東北学院大学工学部卒業。2022年よりBBSakura Networks(株)に所属。現在は同社でモバイルコアとデータプレーンのソフトウェア開発に従事。WIDEプロジェクトメンバー。



深川 祐太

2022年 慶應義塾大学政策・メディア研究科修士課程修了。同年よりNTTコミュニケーションズ(株)に所属。高速ソフトウェアルータKamueeの研究・開発に従事。慶應義塾大学SFC研究所上席所員。WIDEプロジェクトメンバー。



植原 啓介 (正会員)

1995年 電気通信大学電気通信学研究科情報工学大学院単位取得退学 修士。2003年 慶應義塾大学にて博士(政策・メディア)を取得。現在慶應義塾大学環境情報学部教授。インターネット移動体通信に関する研究に従事。IEEE, ACM各会員。WIDEプロジェクトボードメンバー。



中村 修

1990年 慶應義塾大学工学部数理工学科博士課程修了 博士(工学)。1993年 慶應義塾大学環境情報学部助手。大規模広域分散環境、インターネットの研究に従事。現在慶應義塾大学環境情報学部教授。慶應義塾CSIRTチーム長。ACM, ISOC各会員。WIDEプロジェクトボードメンバー。