

エッジコンピューティング時代のサービス運用の課題 ～オペレーター目線で感じる期待と課題～

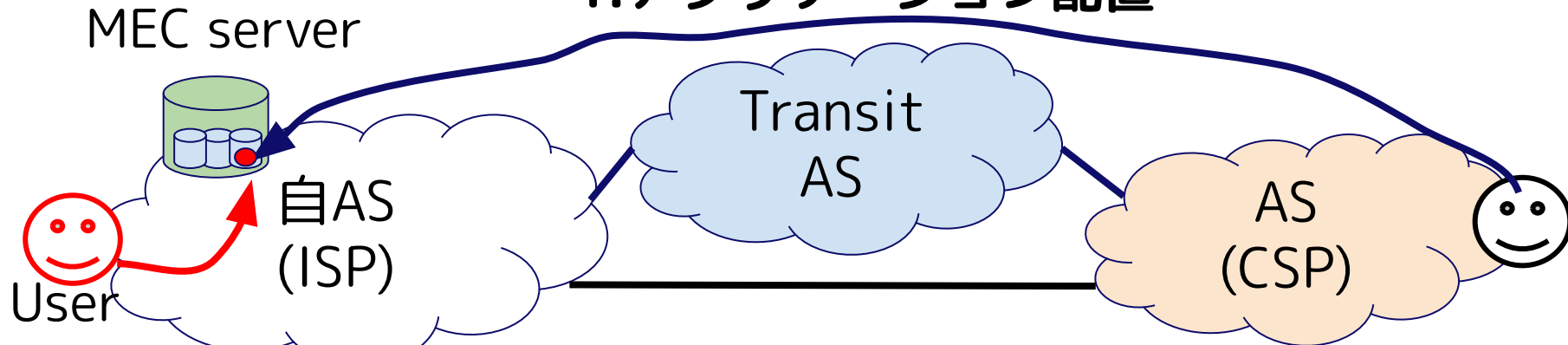
JANOG46 2020/08/27

KDDI総合研究所
宮坂 拓也

MEC? Edge computing?

- ユーザーから“近い”サーバー(MECサーバー)にアプリケーションを配置し、ユーザー～アプリケーション間の低遅延通信や、サーバー側処理の分散化(Distributed computing)を実現

1. アプリケーション配置



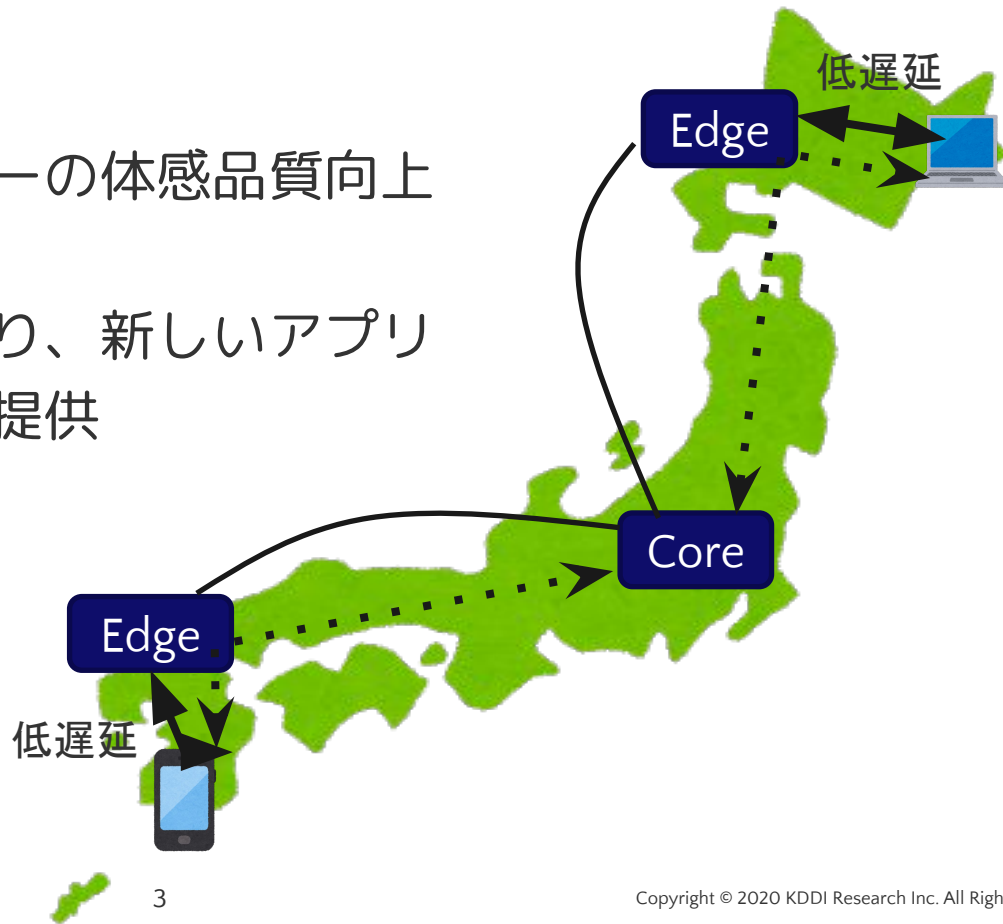
2. 低遅延通信、分散処理

※ISP: Internet Service Provider
※CSP: Contents Service Provider

MEC, Edge Computingに期待すること

1. ユーザー品質向上

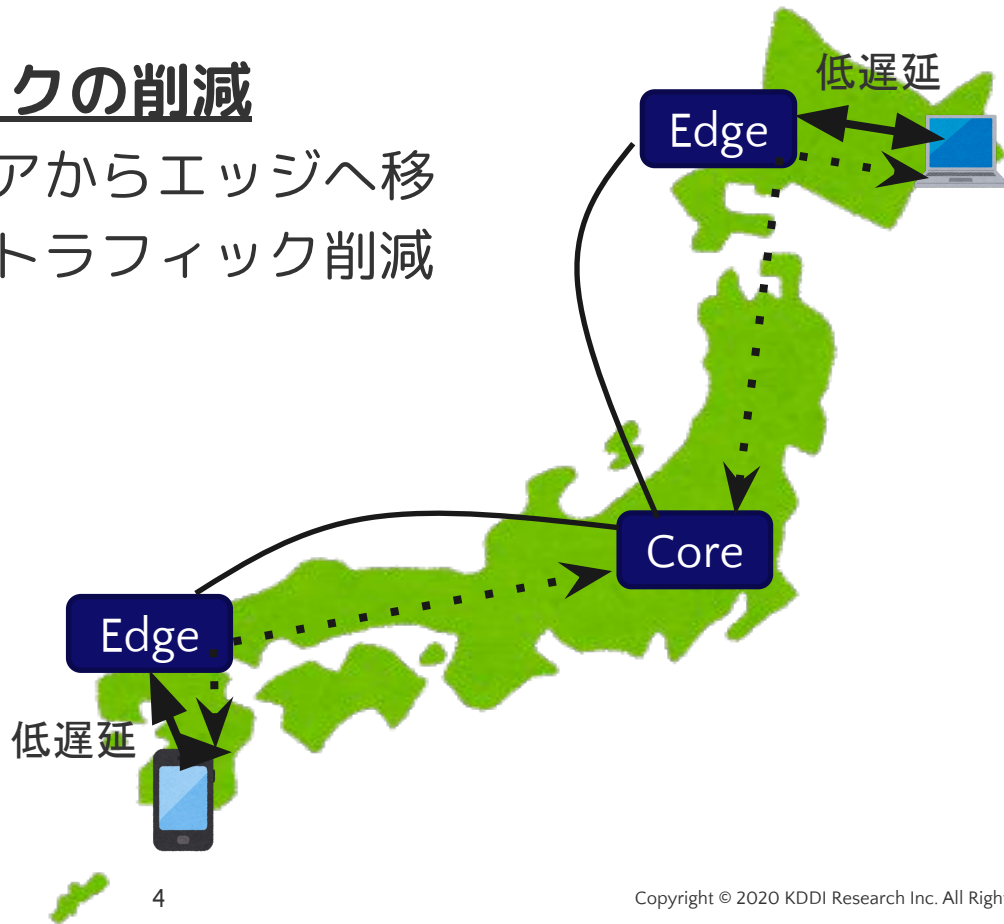
- ▶ 遅延が短くなり、ユーザーの体感品質向上 (“ping値”の低下)
- ▶ エッジサーバー導入により、新しいアプリケーションをユーザーへ提供
 - AR/VR
 - Cloud gaming



MEC, Edge Computingに期待すること

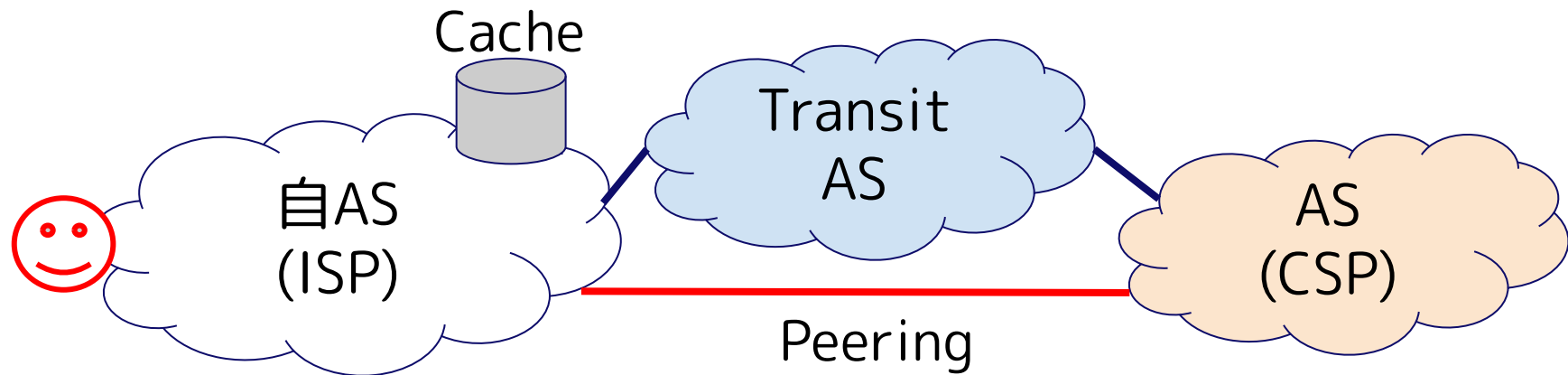
2. バックボーントラフィックの削減

- 通信先インスタンスをコアからエッジへ移行し、コア～エッジ間のトラフィック削減
 - コスト削減



動機的にはPeeringやCacheと同じ

- Peering
 - Transitトラフィック削減、Peer先ASとの通信品質向上
- Cache server
 - Transit/Peerトラフィック削減、Cacheコンテンツの品質向上



オペレータが考えないといけない課題

- Edge computingを実現するためには . . .
 - エッジサーバーの配置
 - エッジサーバーをCSPへ提供
 - エッジサーバーへのユーザーリクエストルーティング

オペレータが考えないといけない課題

- Edge computingを実現するためには・・・
 - エッジサーバーの配置
 - エッジサーバーをCSPへ提供
 - エッジサーバーへのユーザーリクエストルーティング

エッジサーバー配置

- エッジサーバーの物理的配置
 - ファシリティの用意
 - 場所・電力・空調・・・
 - サーバーインフラの用意
 - CPU, Memory, GPU, Network…
 - 特にEdge computingではGPUを用いたユースケースが多いため、どれだけのリソースを用意するか適切に計画する必要がある

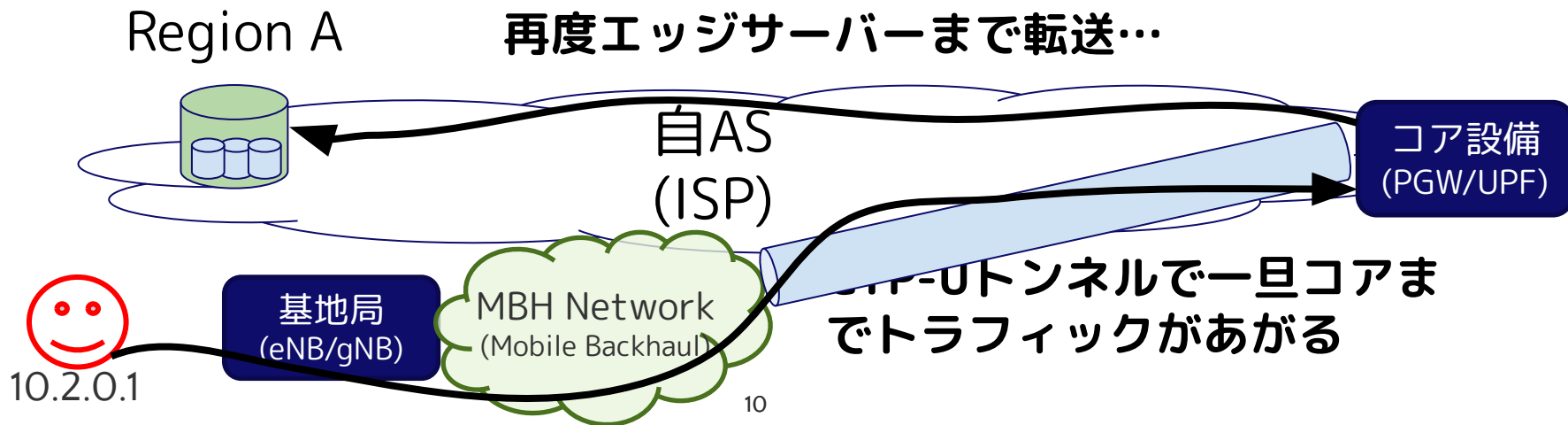
エッジとファシリティとのジレンマ

- エッジはファシリティ的に過酷な環境にあることがある
 - 小さなデータセンターで使える場所が狭い
 - 工場やオフィスの一角かもしれない
 - 携帯電話基地局のような外かもしれない
- そのような環境の場合、電力消費の激しい**高性能なサーバー**を十分に置けない可能性がある・・・



ネットワークを終端できるか？

- ユーザーのIPパケットがトンネルによって、終端点がコアまで行ってしまふこともあるので要注意 (特にモバイル)
 - エッジサーバーを用いることでかえって遅延増大することもある
 - 詳細：<https://www.janog.gr.jp/meeting/janog45/program/mec>



オペレータが考えないといけない課題

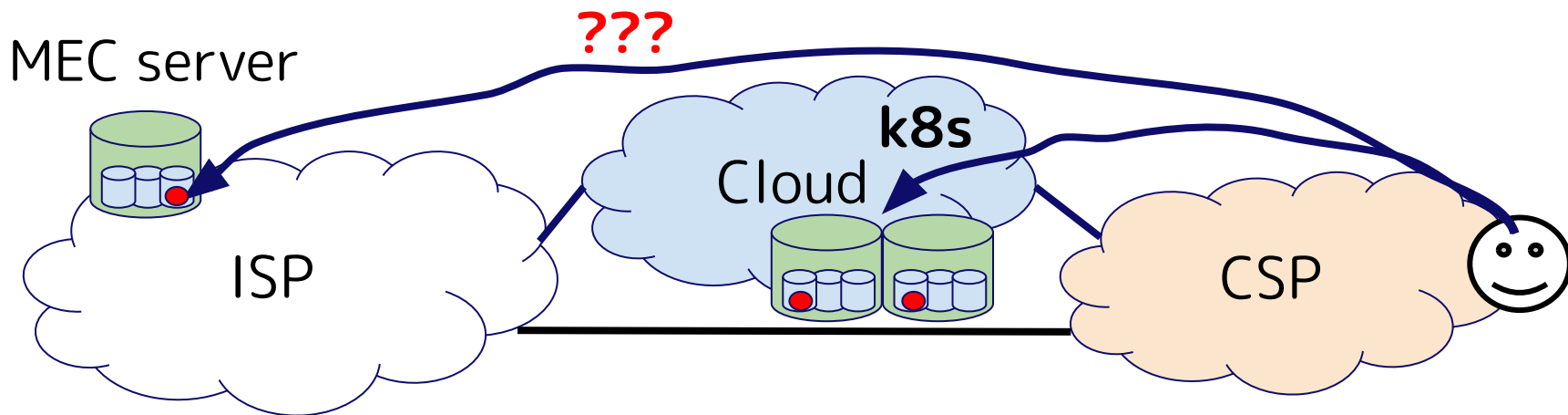
- Edge computingを実現するためには・・・
 - エッジサーバーの配置
 - エッジサーバーをCSPへ提供
 - エッジサーバーへのユーザーリクエストルーティング

エッジサーバーの提供

- エッジサーバーをCSPに提供
 - CSPが自由にアプリケーションインスタンスを起動
 - どのような形態を許可するか決めないといけない
 - Bare metal?
 - VM?
 - Container?
 - いかにか、CSPがやりやすい形で提供できるかが肝
 - 通常利用しているクラウドとのハイブリッド構成も

Hybrid Cloud

- Hybrid Cloud: CSPの観点では、すでにクラウドなどでサービスを運用しており、その一部をISP内のエッジサーバーへ張り出す
 - その場合、「ISPは、CSPに運用しやすい形(例:k8s)でエッジサーバーをCSPへ提供できるか？」がとても重要な問題となってくる

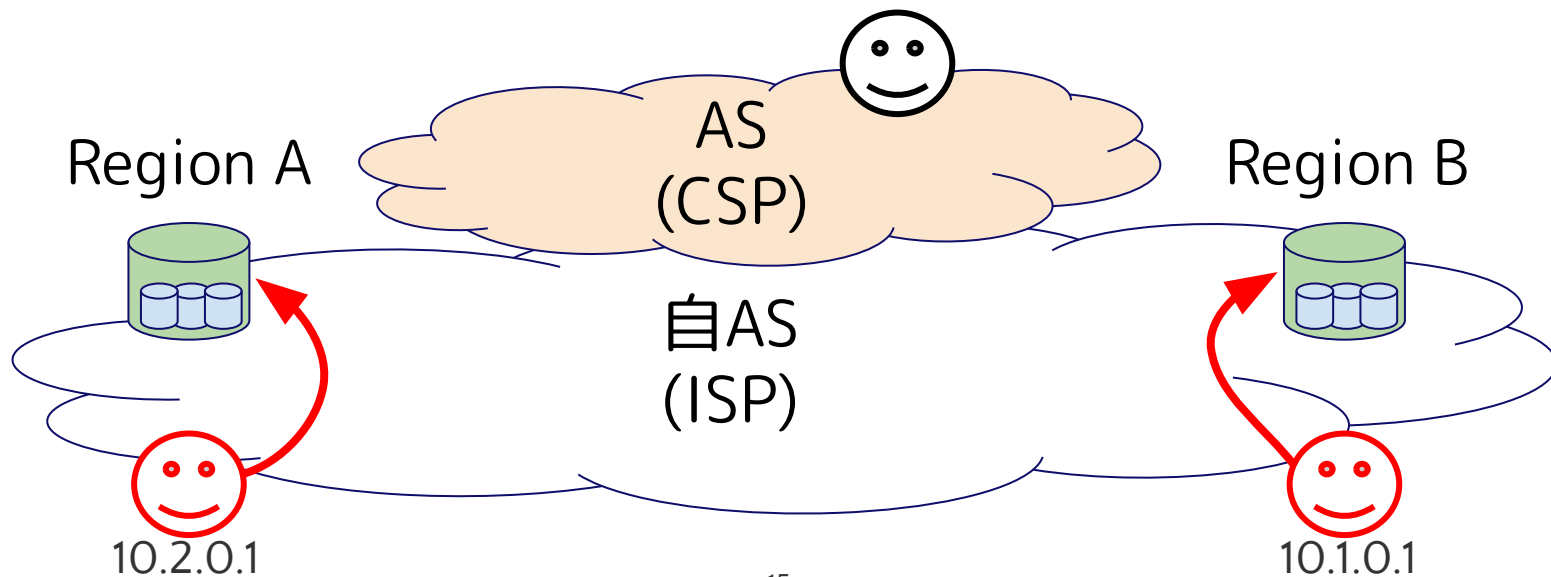


オペレータが考えないといけない課題

- Edge computingを実現するためには . . .
 - エッジサーバーの配置
 - エッジサーバーをCSPへ提供
 - エッジサーバーへのユーザーリクエストルーティング

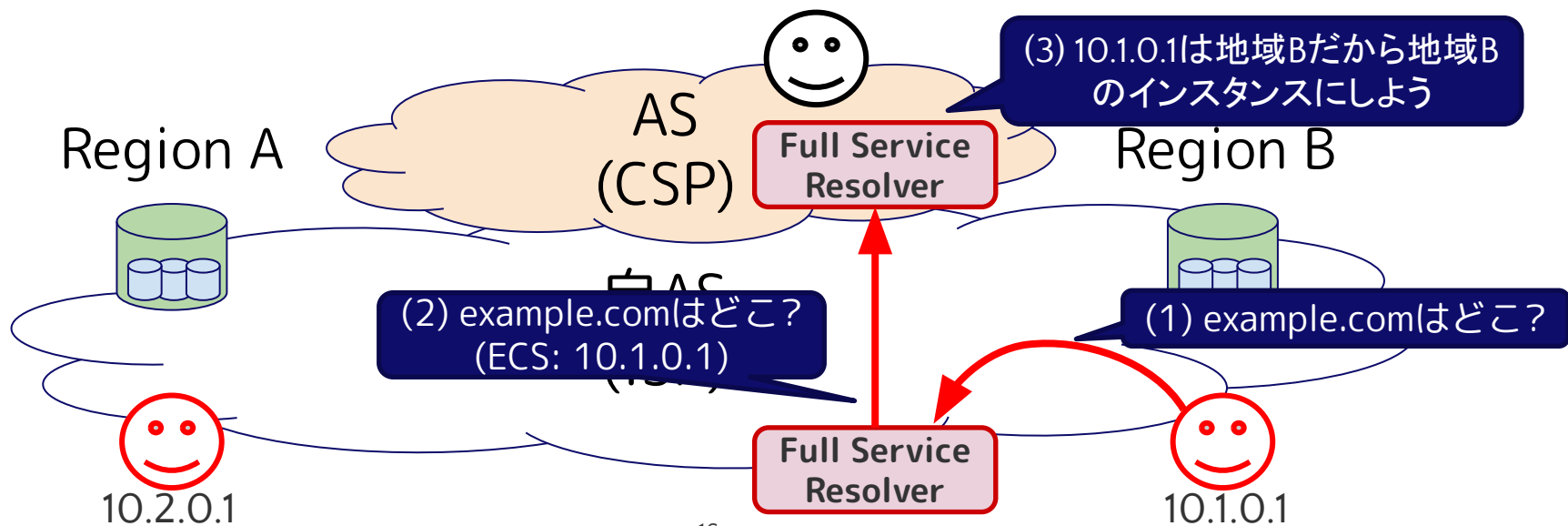
ユーザーリクエストのルーティング

- ユーザーからのアプリケーション接続要求を、そのユーザーに適切なMECサーバーへ振り分ける (例：DNS, IP anycast, Redirect)
 - **ISPとCSPの「連携」が必要になることが多い**



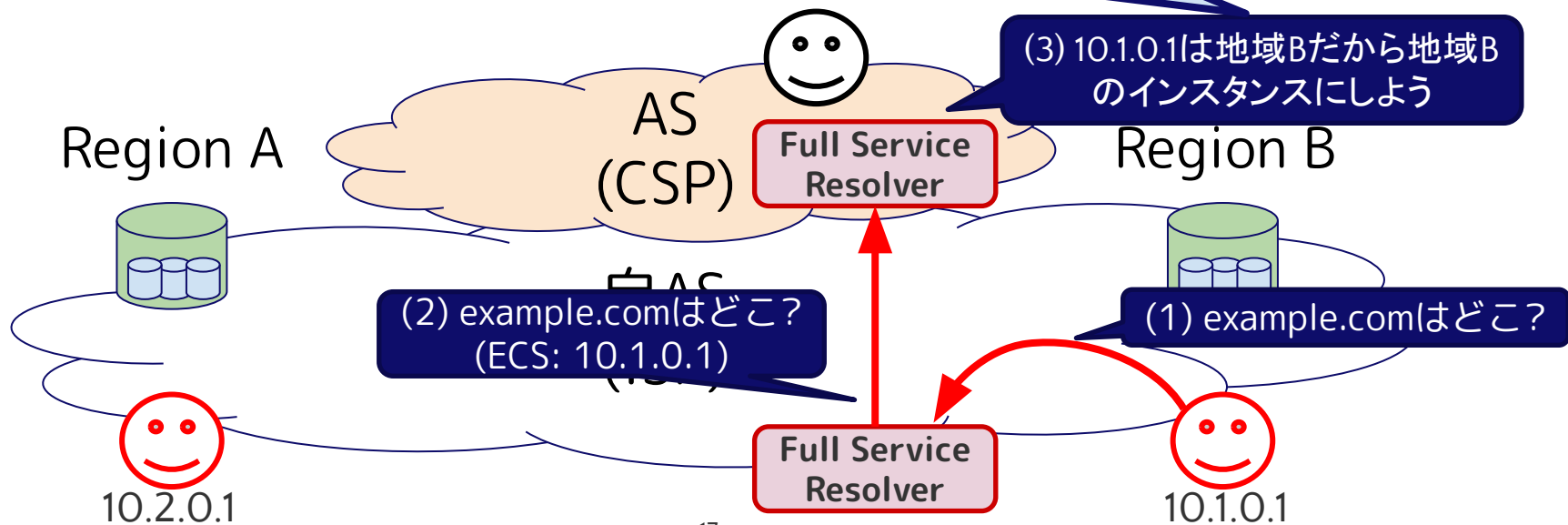
例：DNSによるRouting

- CSPの権威サーバーにおいて、(1)DNSクエリの送信元フルリゾルバーのIPアドレス、(2)ECSによって通知されたクライアントのIPプレフィックスより、地域を特定しその地域インスタンスのIPアドレスを返す



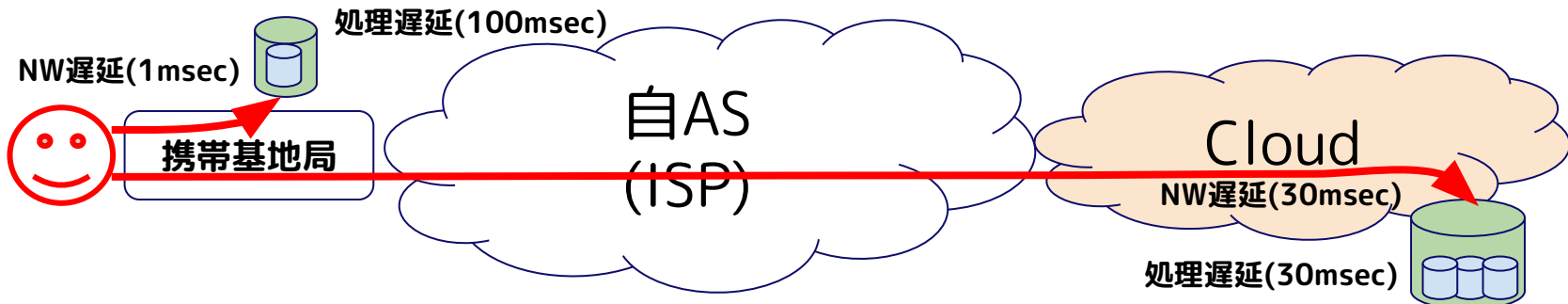
例：DNSによるRouting

- CSPのIPアドレスよりこの方法では、あらかじめISPからCSPに対して、「地域毎のIPアドレスリスト」的なものを何らかの方法で渡す必要がある



必ずしもネットワーク的に近い=最適ではない

(1) エッジサーバー



(2) クラウド

- エッジサーバーは環境により、マシンパワーが不足していることも
- そのような場合、クラウドなどエッジサーバーを利用しないほうが、ユーザー体感品質が向上することも往々にして存在する
 - End-to-end 遅延 = ネットワーク遅延 + サーバーでの処理遅延

まとめ

- Edge computingに対する期待
 - ユーザー品質向上
 - バックボーントラフィックの削減
- Edge computing提供のための課題
 - エッジサーバーを配置できる？
 - エッジサーバーをCSPに提供できる？
 - エッジサーバーへユーザーリクエストルーティングできる？