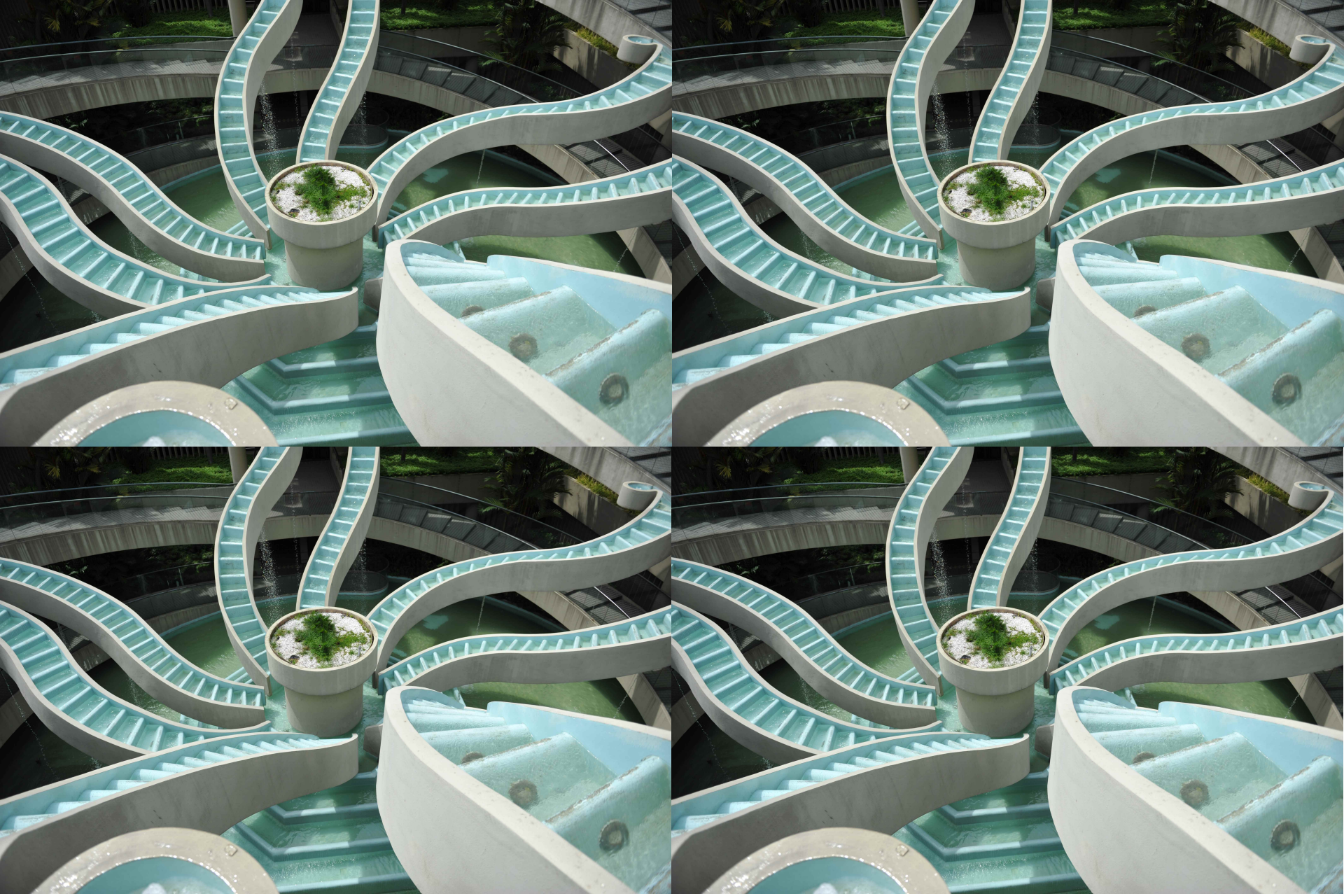
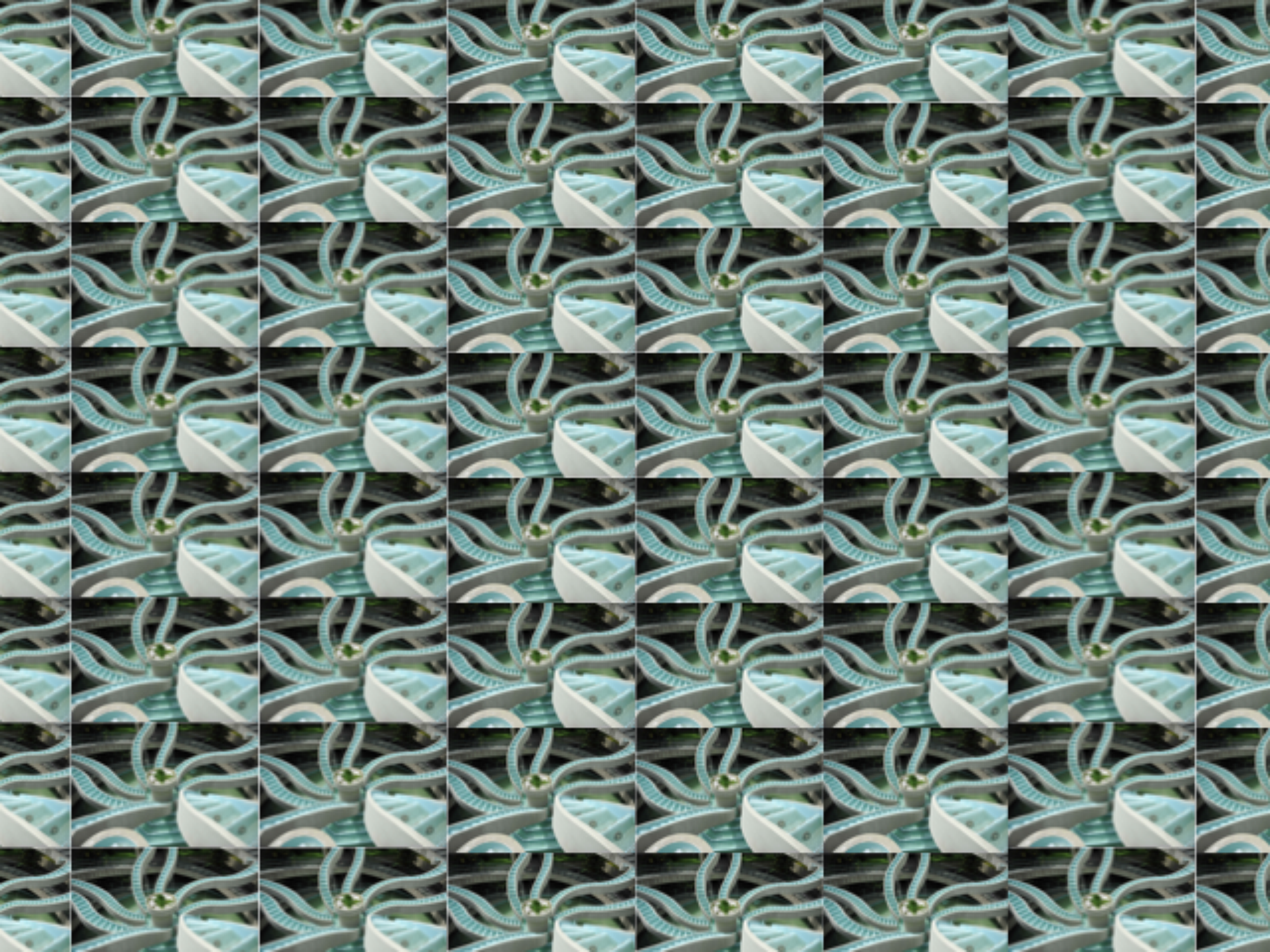




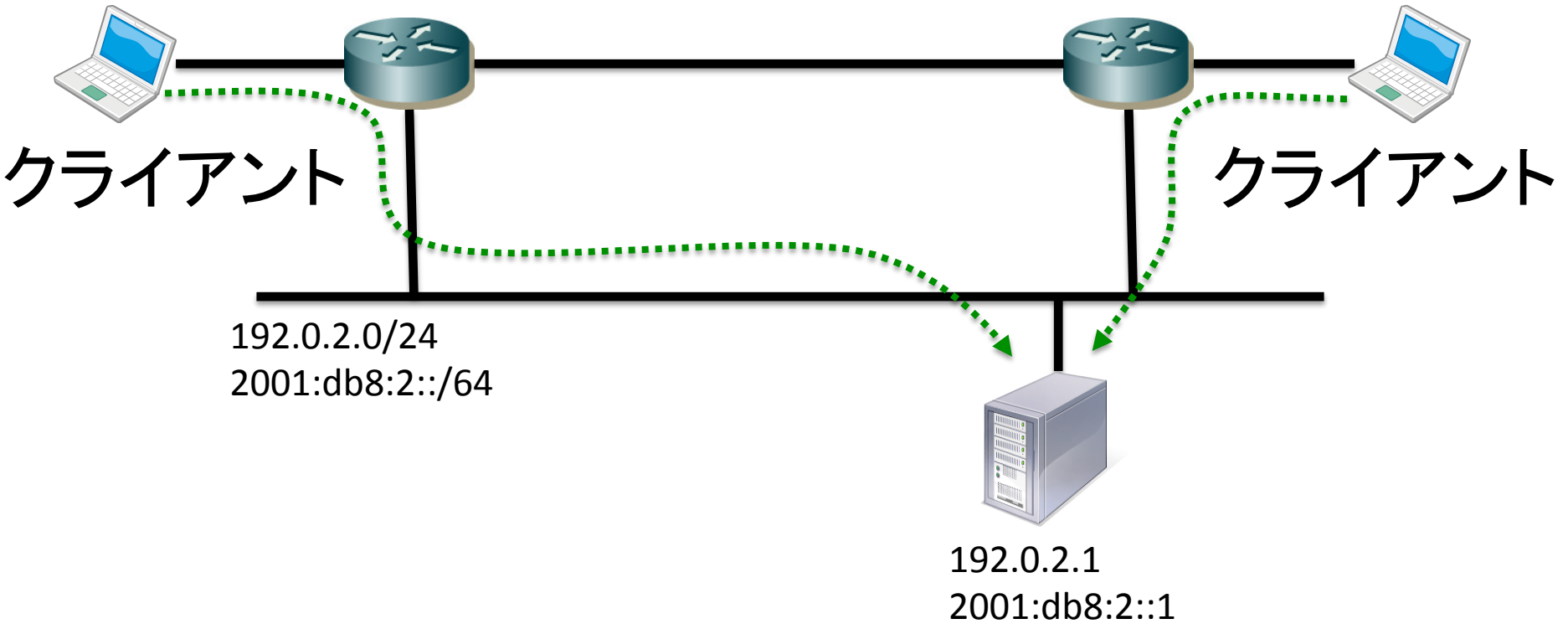
頑張れIP anycast

Matsuzaki ‘maz’ Yoshinobu
<maz@iij.ad.jp>

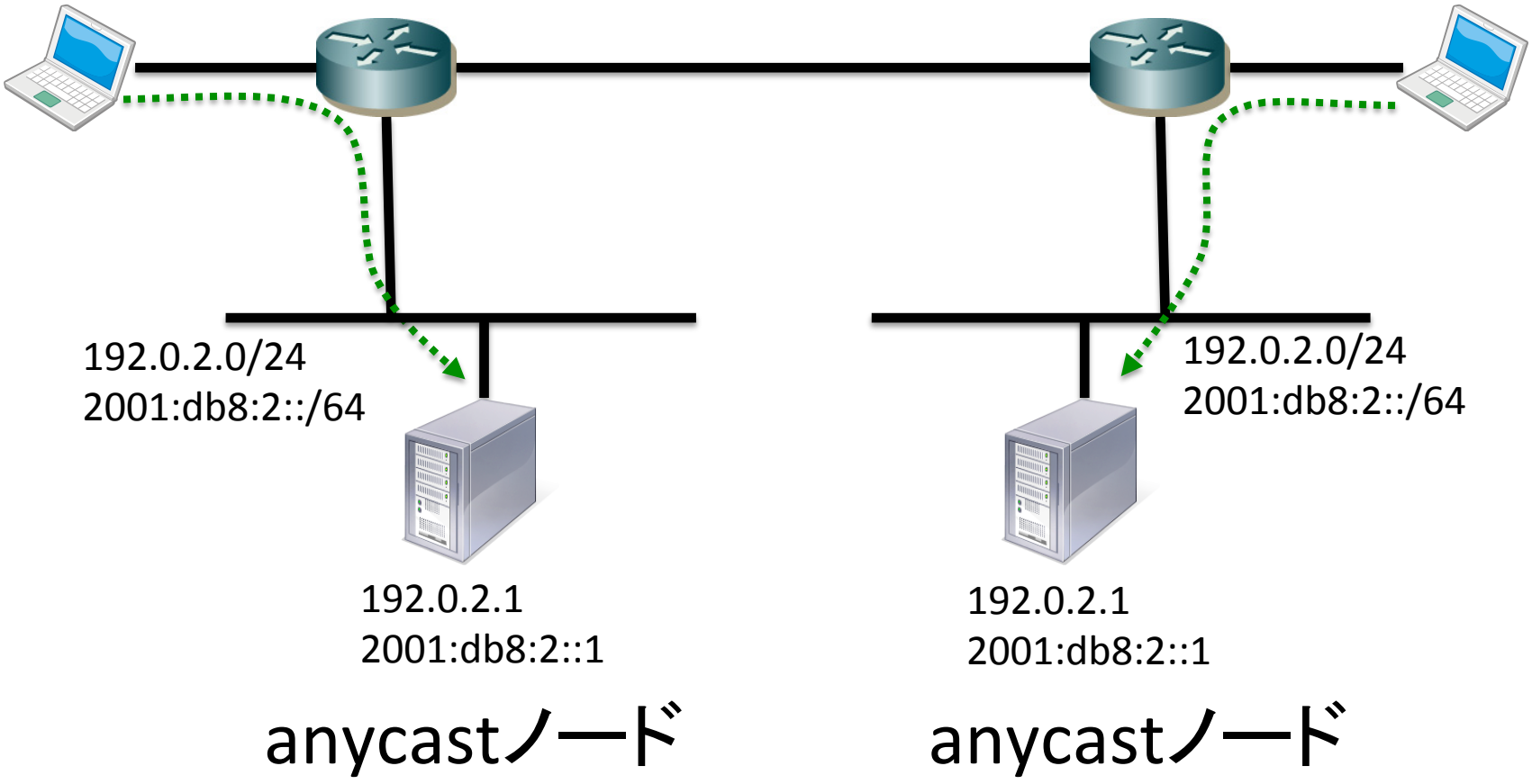
IP anycast

- 主にサーバ側で利用する技術
- 実は単なるunicast
  - 複数箇所に同じIPネットワーク
  - でも、ルータは単に宛先に投げてるだけ
  - anycastは状態だと思うのが良いかも
- ユーザからはanycastのノード数は分からない
  - 1以上のノードが稼働していればサービスは可能

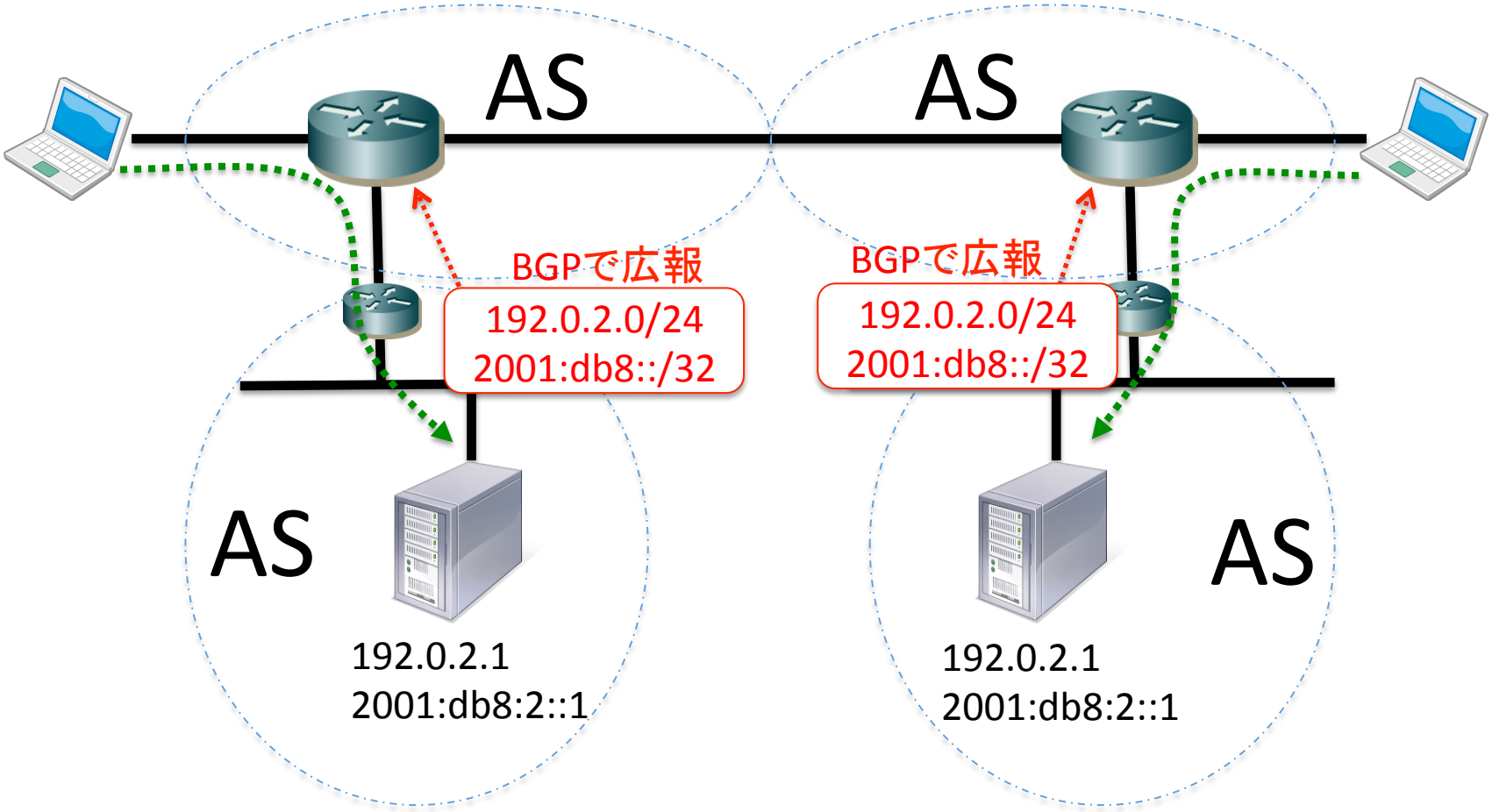
クライアントとサーバ



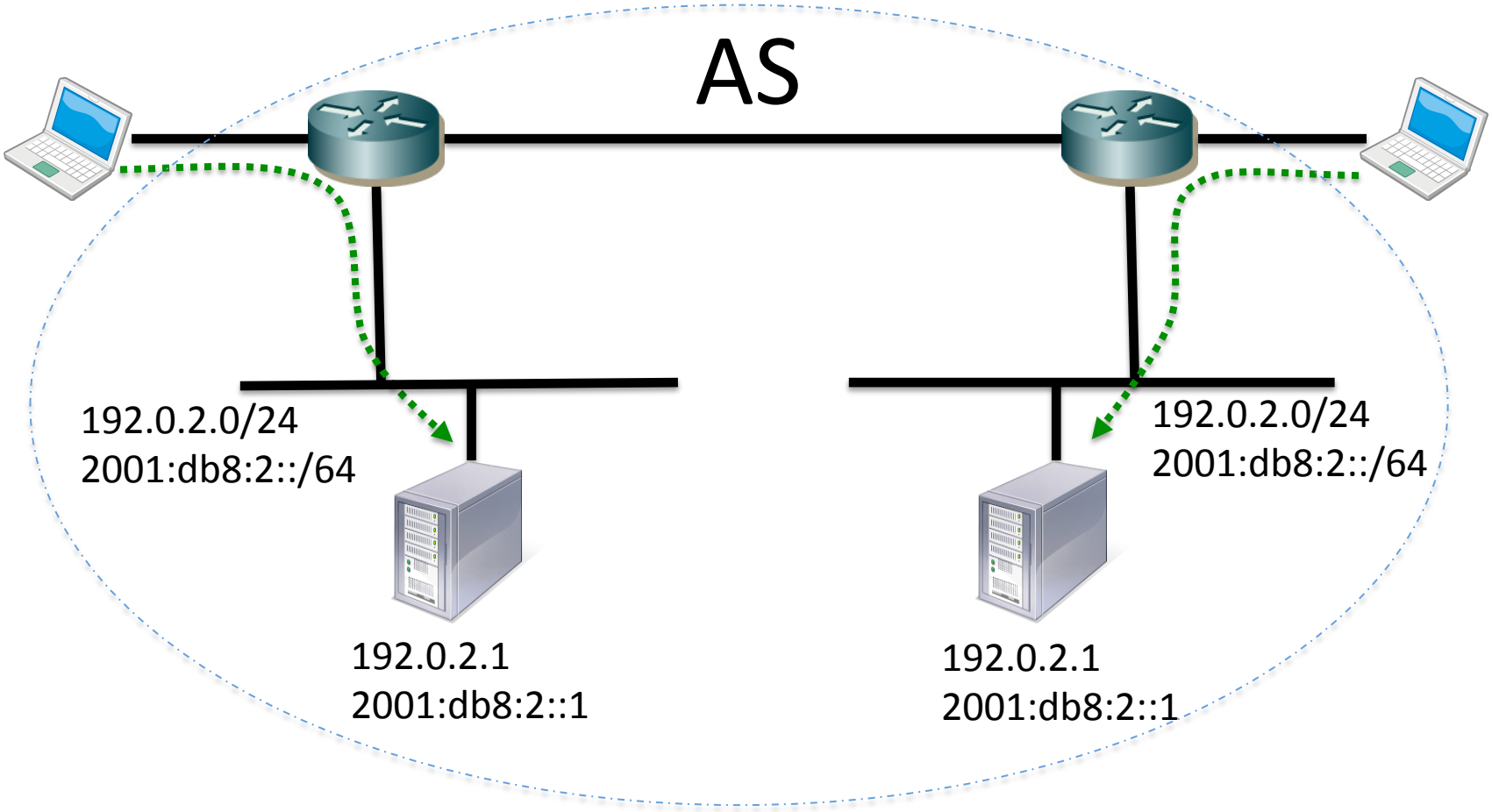
複製したら、ほらIP anycast



AS間だと、BGPで制御



AS内だと適当に制御


AS
192.0.2.0/24
2001:db8:2::/64
192.0.2.1
2001:db8:2::1
192.0.2.0/24
2001:db8:2::/64
192.0.2.1
2001:db8:2::1

AS内でのanycast制御

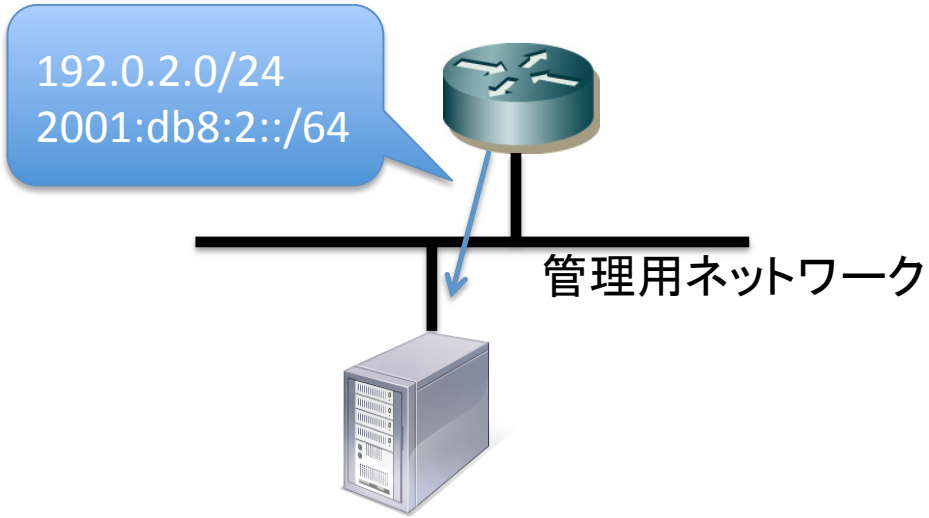
- unicastの経路制御と同じ手法
  - 矛盾さえ無ければ、何だって使える
  - connected
  - static
  - OSPF、IS-IS、BGPとかとか
- prefix長も好きにして良い
  - /32とか/128でもいいし、/24とか/64でも大丈夫

anycast用アドレスのBGP経路広報

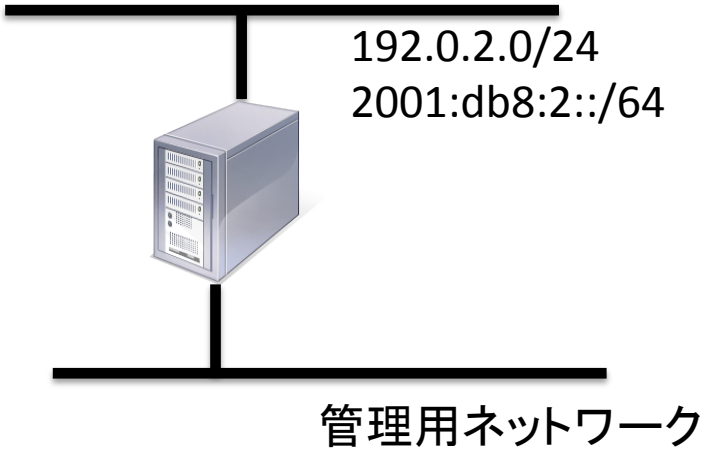
- 大きなprefixに含めて広報
  - PAブロックとか
- 専用のprefixだったら、そのまま広報
  - /24とか、/48とか

普通は管理用のIPアドレスも付けるよ

- anycastはloopback運用
  - 別途経路制御が必要
    - static or dynamic



- 別インタフェースで実装
  - インタフェースが複数必要



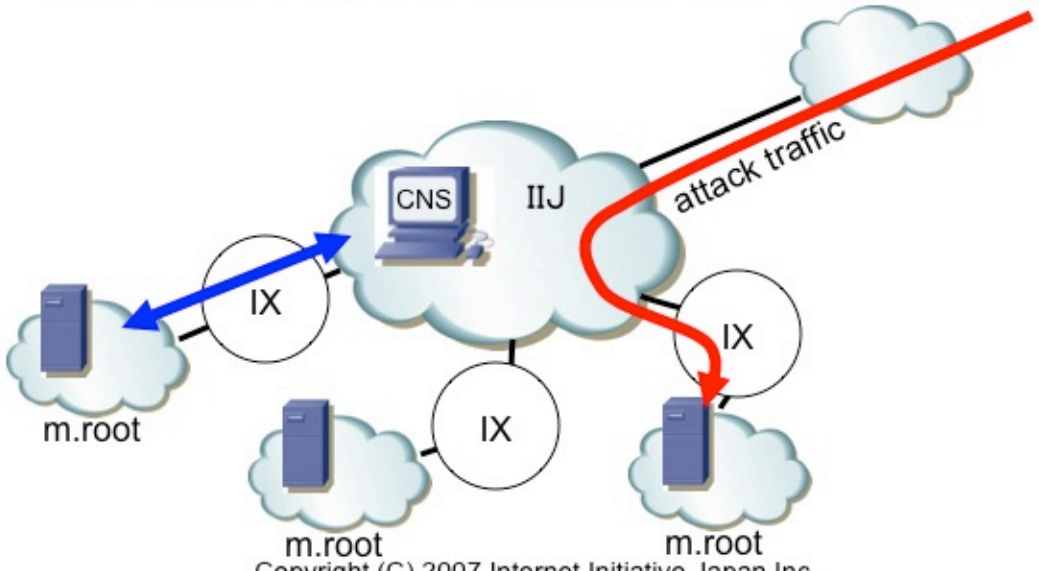
IP anycastの嬉しさ

- 一つのIPアドレスでサーバを地域分散できる
  - 遅延の軽減
  - 負荷分散
- 攻撃を局所化できる
  - どのノードを利用するかユーザ側で制御できない
  - 他のノードを直接攻撃するのは難しい
- anycastノードをそれぞれ独立して運用できる
  - 障害の連鎖とかが少ない

2007/02、rootへの攻撃時

during the attack

- IIJ transited attack traffic as well...
  - IIJ's cache server selected the other site.



17

こんな事にもanycast使ってました

- 全ルータのloopbackに同じIPアドレスを追加
  - connected経路に見えるので、IGPでは広報せず
  - NTPサーバとして参照すると直近ルータが答える
  - tracerouteすると流入点のルータが応答
  - 参照用のNTPサーバを専用に用意して縮退
- ユーザの参照用DNS
  - 直近のノードが応答する
  - ばらまきすぎて効率が悪くなったので縮退

実際のIP anycast – 権威DNS

- root DNS
  - http://www.root-servers.org/
- JP DNS
  - http://www.dns.jp/
  - d.dns.jp の事例
    - http://www.iij.ad.jp/company/development/tech/activities/ddnsjp/

実際のIP anycast – キャッシュDNS

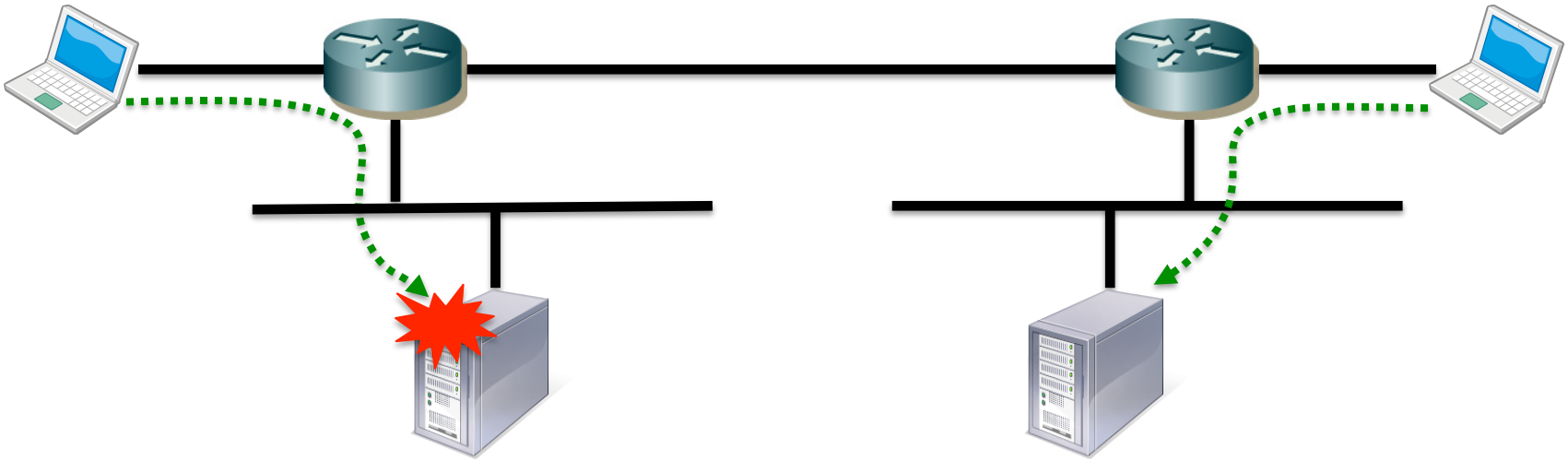
- google public DNS
  - https://developers.google.com/speed/public-dns/
- OpenDNS
  - http://www.opendns.com/

実際のIP anycast – CDN

- CloudFlare
  - https://www.cloudflare.com/
- Microsoft Azure CDN
  - http://azure.microsoft.com/ja-jp/services/cdn/
  - [janog:12473] からのスレッドも参照

ノード障害とブラックホール

- サービス障害時、経路を迂回させるか
  - 該当経路の広報停止や優先度の変更



IP anycast実装の難しさ

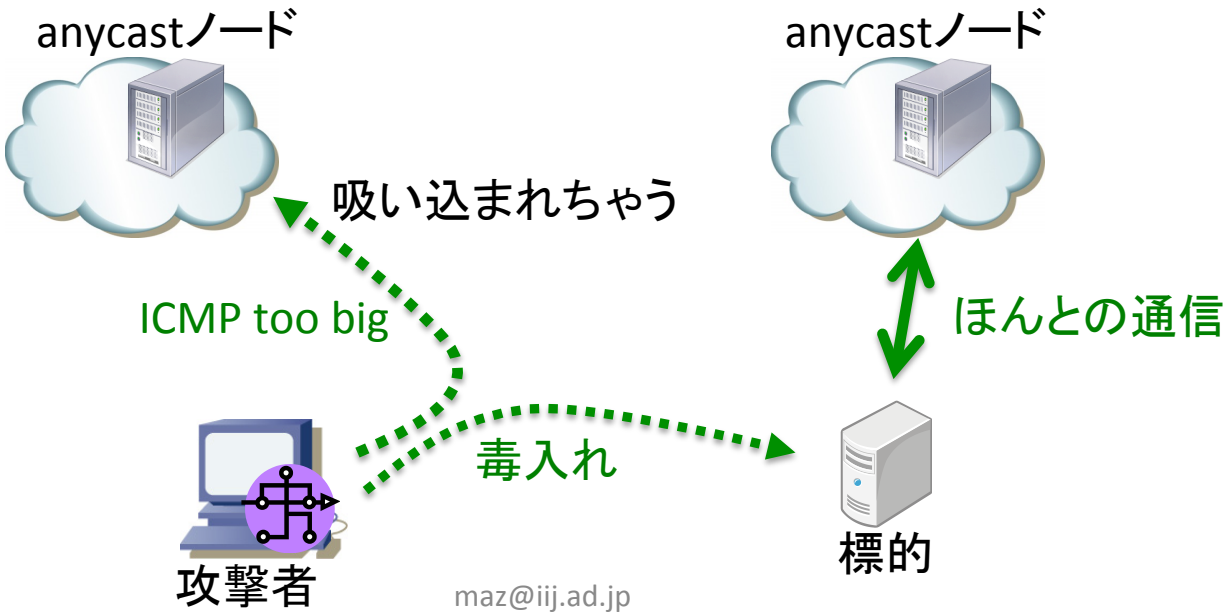
- 監視
  - 生死、コンテンツ、サービスの一貫性
- サービスと経路制御の連携
  - 経路制御必須
  - サービス連携が無ければブラックホール化も
- ノード間での負荷分散
  - ノード縮退時の再分散も難しい
  →入りのトラヒック制御が難しいのと同じ
- トラブルシュート

IP anycastの制限事項

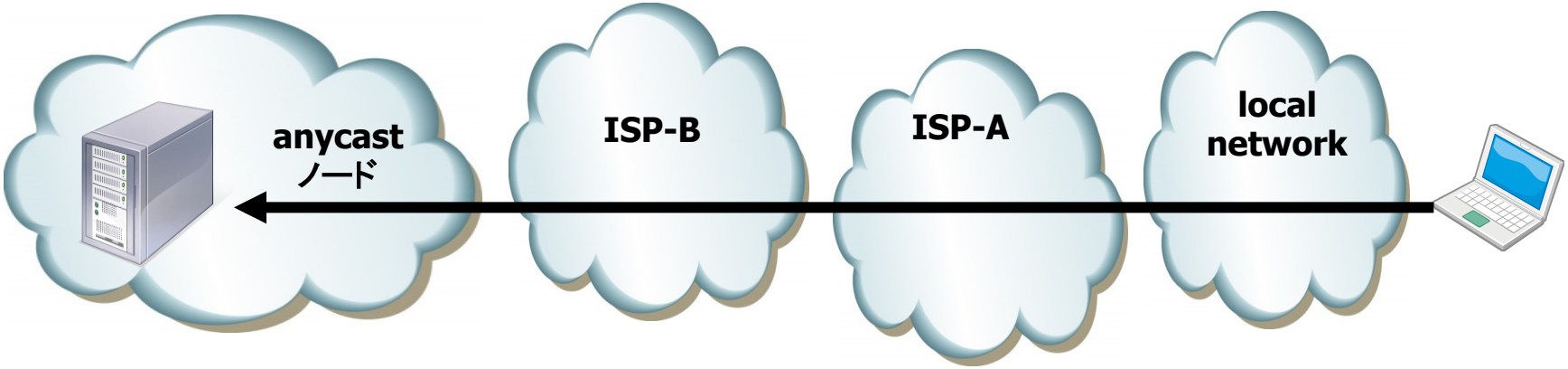
- 同じクライアントからの通信でも、異なるanycastノードにパケットが届くかも
  - 経路変動やmultipathでの負荷分散
  - 双方向通信が継続できないかも
- 途中ルータからは異なるノードに向かうかも
  - Path MTU Discoveryが動かないかも
    - anycastノード -> クライアント方向で大きなパケットの際
  - その他、ICMPエラーもきちんと受け取れないかも

IP anycastとfragment便乗攻撃

- 応答サイズの調整が必須の場合、ICMPパケットを標的が参照しているanycastノードに届ける必要がある
  - IP anycastされていると、これが難しい

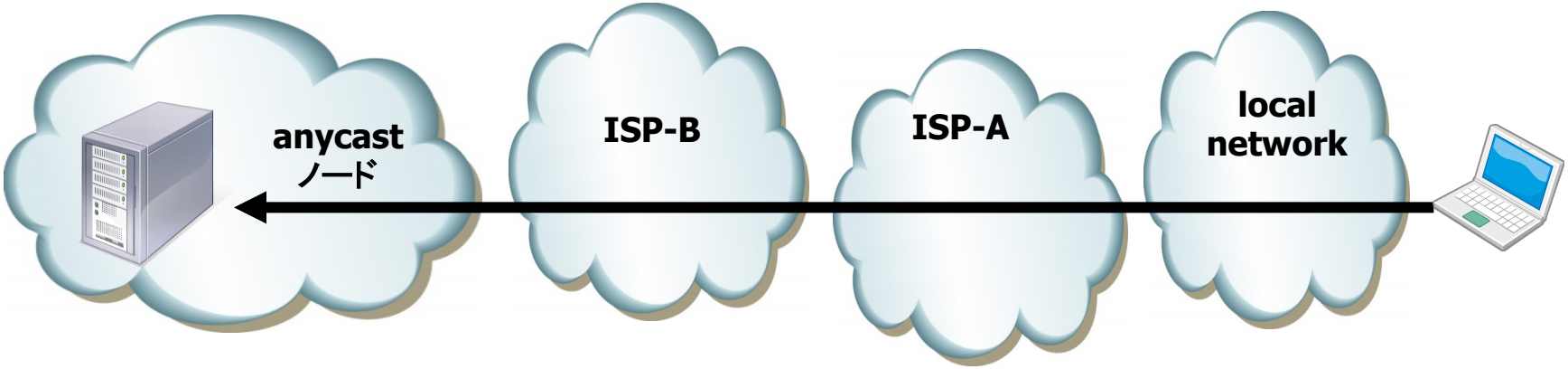


障害切り分け



- このそれぞれにIP anycast故の複雑さが絡む
  - ノード障害
  - 経路障害
  - パケットフィルタ

ノード特定



- traceroute
- アプリケーションで応答
  - id.server (l-rootの場合 RFC7108)
- まだ標準手法はなさそう

障害連絡

- ユーザ側ではサービスがIP anycastかどうか分からない
  - URLや対象IPアドレスで障害申告する
    - anycastノードとしては一意ではない
- 障害受付側でも、サービスやネットワークに関する知識が必要
  - 手元ではうまく動く場合があり、申告したユーザ固有の問題に見えることも

聞いてみたい

- IP anycast利用事例
  - こんな事に使ってる！
- IP anycast運用＆導入悩み
  - 管理とかどうしてる？
  - 心配事がある？
- IP anycast利用者の立場
  - 近い事は良いことだ？
  - 困ったときに困る？

おわり

anycastの恐怖

- ノードを追加したけど、経路が広報されてなくて、anycastに参加してなかった
- BGPコミュニティ(no-export)で制御していたら、経路が届いていないASがあった
- 投入したノードのサービスがanycastアドレスをlistenしてなかった
- 運用者がanycastを理解してなくて、トラブル発生。トラブルシュートできないおまけ付き

監視の課題

- リモートからは、どのanycastノードを監視してるか分からない
  - サービスしてるのはanycast用アドレス
  - リモートから監視できるのは管理用IPアドレス

- anycastノード自身で監視？
  - プロセス監視
  - ローカルで接続テスト

予備ノード

- いざって時のためのバックアップ
  - 正副に加えて予備系での冗長化
- 経路の優先度を落として、日頃見えない
  - 攻撃者からも見えない
- 困ったら簡単な設定変更でサービス開始
  - 経路優先度の変更
  - 他ノードの停止