

How to Support Multi-AZs in NFV Services

Kento KAWAKAMI, Verda Network Development Team, LINE Corporation

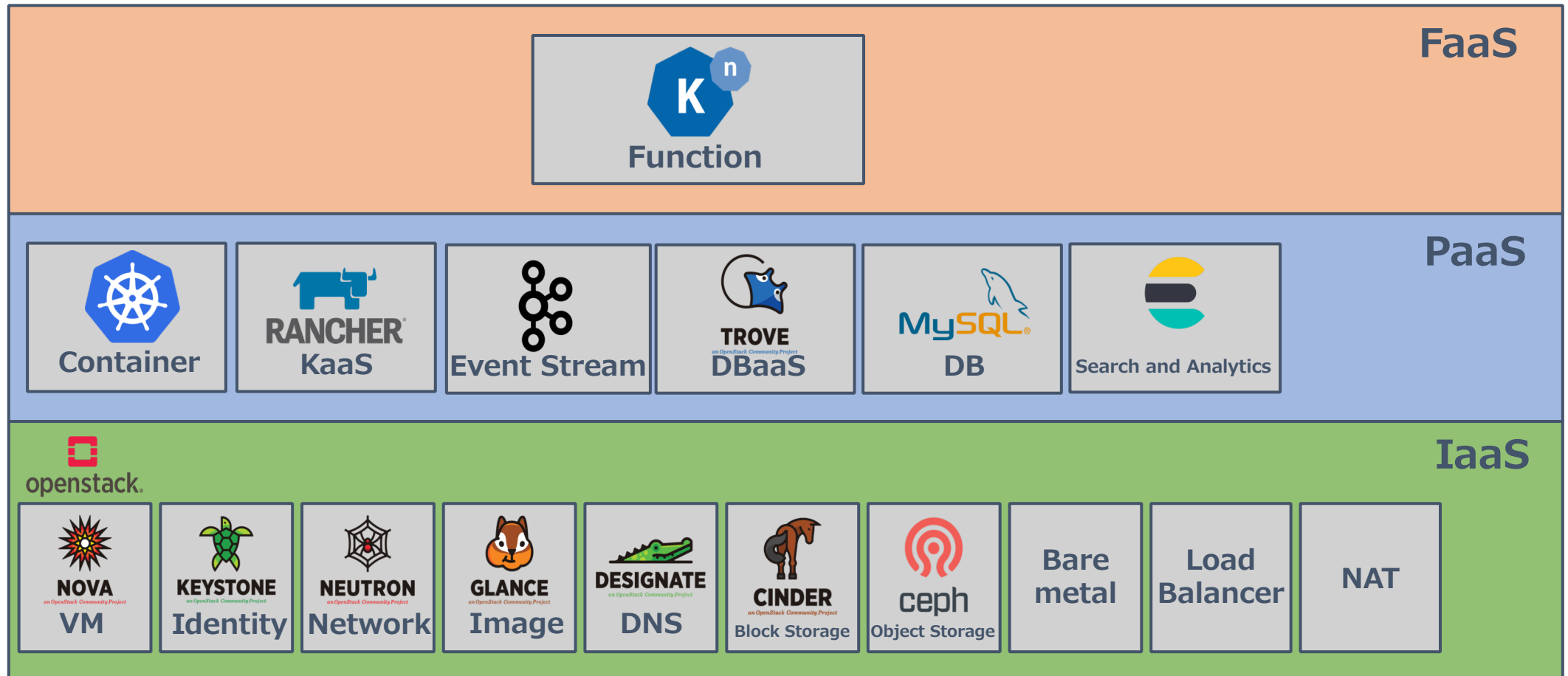
ABOUT ME

- プロフィール
 - 名前：川上 けんと (Kento KAWAKAMi)
- LINE Corporation
 - ITSC Network Development Team
- 職歴
 - 2019年～：LINE (新卒入社)
DNS Serviceの設計、運用、開発
NAT Service, LB Service, Overlay NWの設計、運用、開発

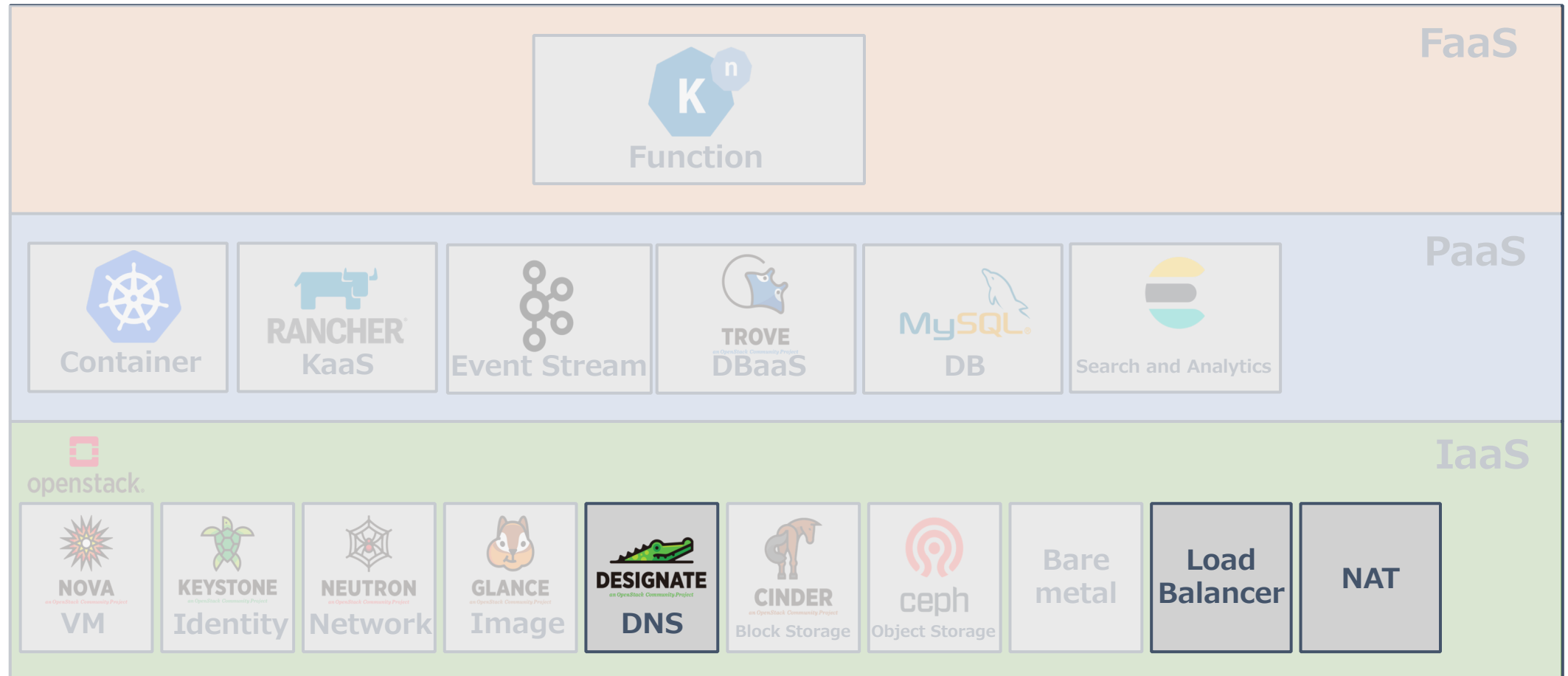
Agenda

- Our NFV Services
- Multi-AZs Support Policy
- How to support Multi-AZs in our NFVs
- Conclusion

Our Services



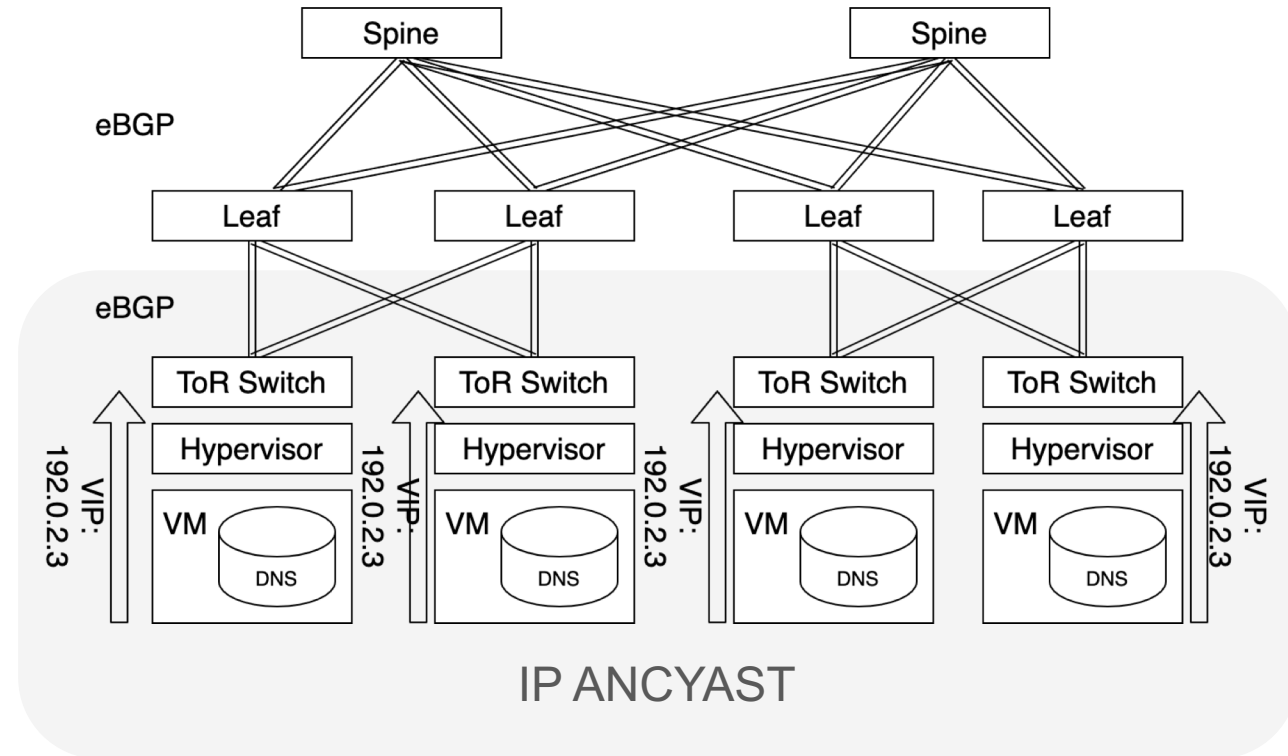
Our Services: NFV



Our NFV Services: DNS

DNS

- DNS CacheとDNS Authを提供
- IP ANYCASTを用いて冗長化



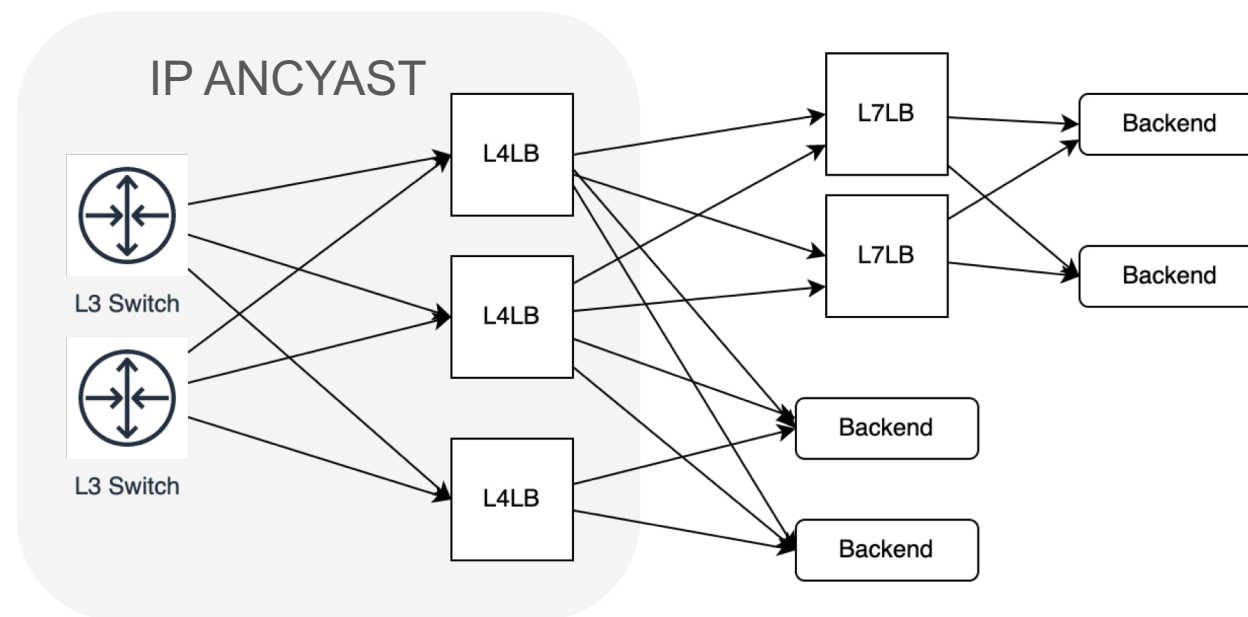
Our NFV Services: Load Balancer

DNS

- DNS CacheとDNS Authを提供
- IP ANYCASTを用いて冗長化

Load Balancer

- L4LBとL7LBを提供
- L4LBはIP ANYCASTを用いて冗長化



Our NFV Services: NAT

DNS

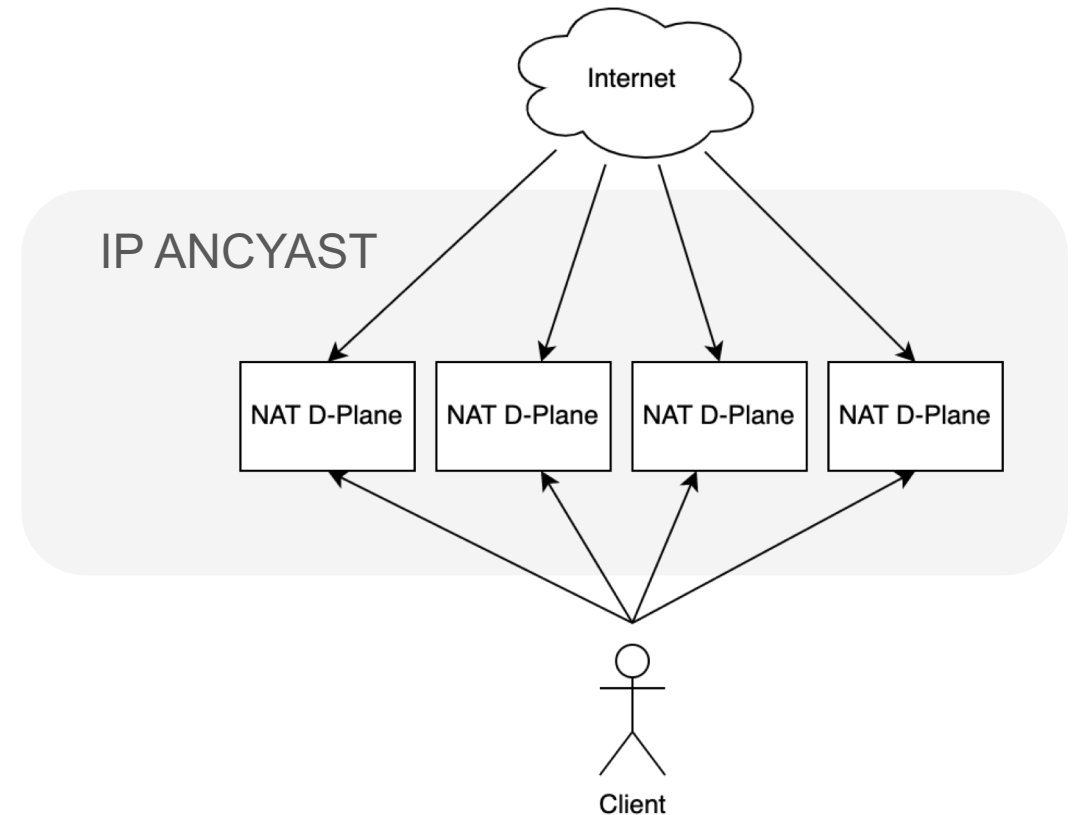
- DNS CacheとDNS Authを提供
- IP ANYCASTを用いて冗長化

Load Balancer

- L4LBとL7LBを提供
- L4LBはIP ANYCASTを用いて冗長化

NAT

- Internetへ向けてのNATサービスを提供
- NATのD-PlaneはIP ANYCASTを用いて分散



Multi-AZs Support Policy in NFV Services

- AZを意識する事なくサービスを利用可能にする
 - Defaultでサービスを利用する場合にはAZが利用される状態にする
 - オプトインでは基本的に利用されない
- 既存のユーザに対してもAZの提供を行う
 - 新規のユーザのみではAZの利用が広がらない
 - 既存のD-Planeをマイグレーションしユーザに提供する

Our NFV Services

DNS

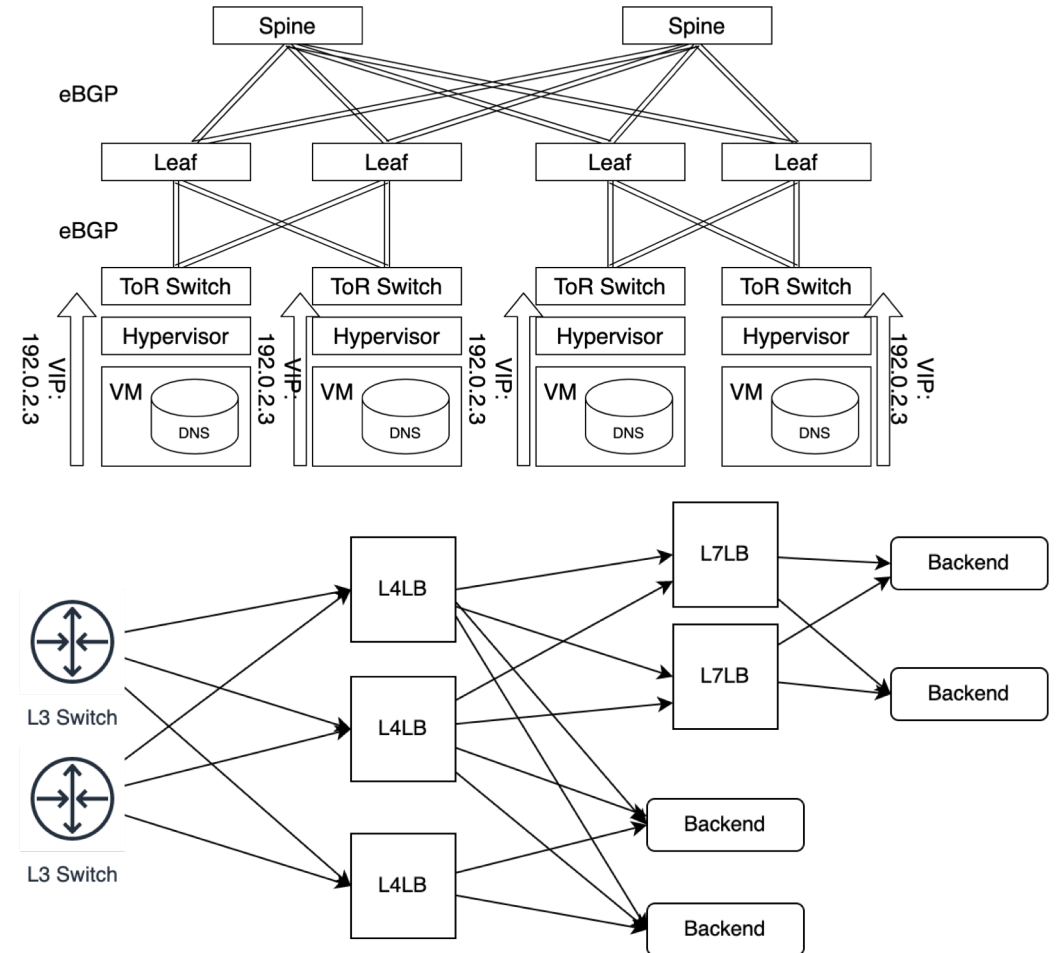
- DNS CacheとDNS Authを提供
- IP ANYCASTを用いて冗長化

Load Balancer

- L4LBとL7LBを提供
- L4LBはIP ANYCASTを用いて冗長化

NAT

- Internetへ向けてのNATサービスを提供
- NATのD-PlaneはIP ANYCASTを用いて分散



Our NFV Services

DNS

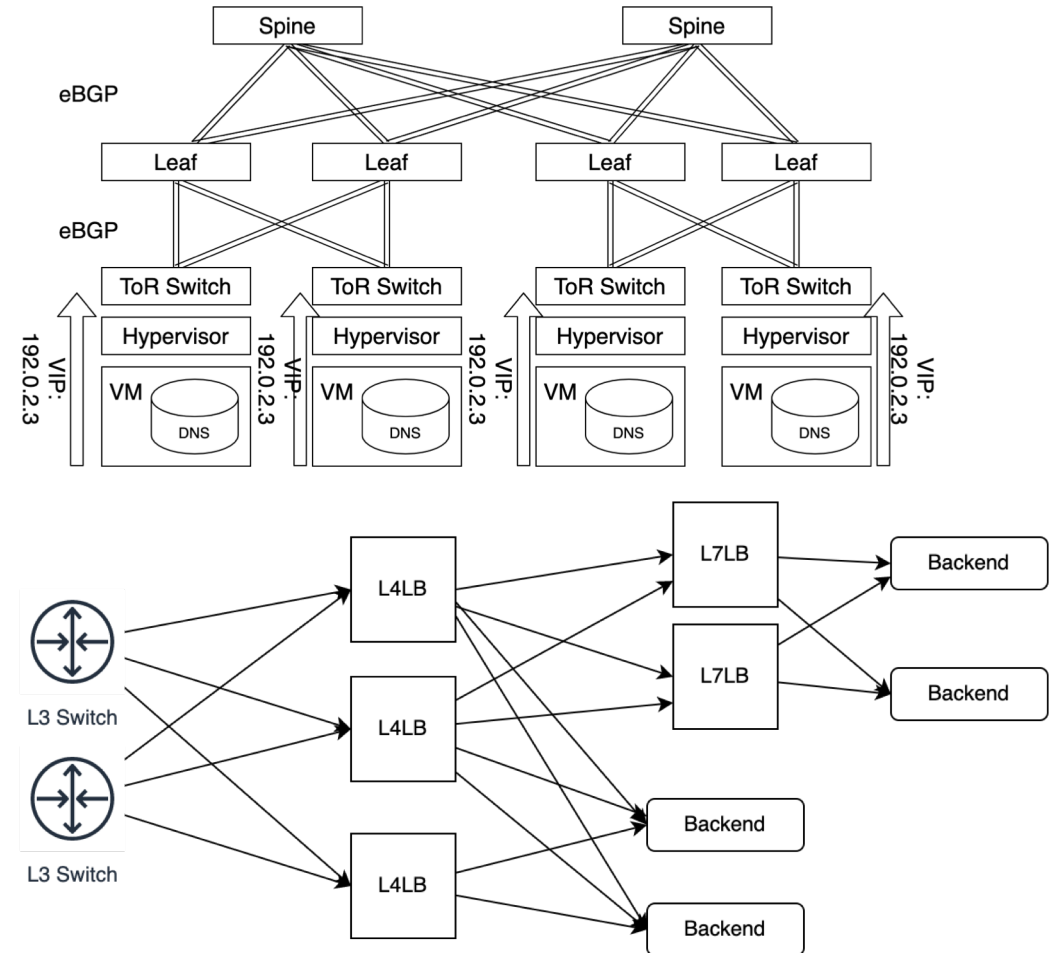
- DNS CacheとDNS Authを提供
- IP ANYCASTを用いて冗長化

Load Balancer

- L4LBとL7LBを提供
- L4LBはIP ANYCASTを用いて冗長化

NAT

- Internetへ向けてのNATサービスを提供
- NATのD-PlaneはIP ANYCASTを用いて分散



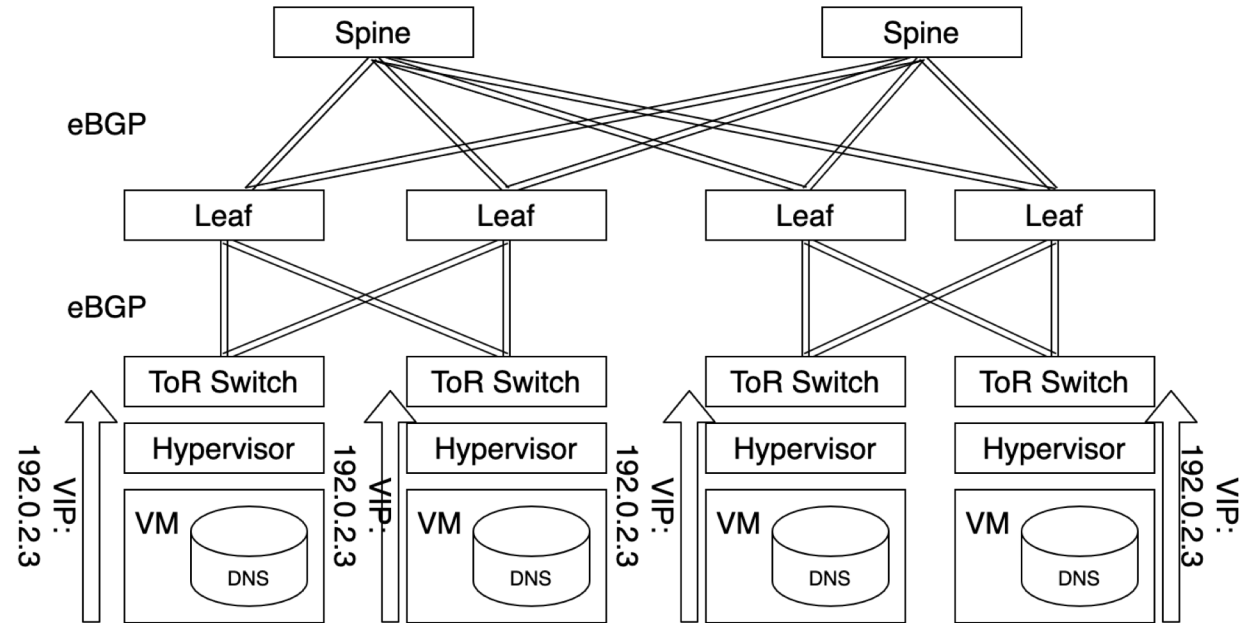
DNS: Architecture

VMで構築

- DNS用のサーバは他のVMと全く同じ構成
- 通常のVMと同じようにHVにスケジューリング

BGPで経路広報

- Full L3のネットワーク構成
- VMからHVを経由してDNSのVIPを広報
- VMの中でBGP Daemonが動作



Conclusion

subtitle

- DNSの冗長化にIP ANYCASTを用いる事で多くのメリットが存在する
 - スケーラビリティの向上
 - CLOS上にVMの追加を行うだけでキャパシティが増強可能
 - メンテナンス性の向上
 - VMから経路のコントロールが可能
 - VIPのポータビリティが向上
- IP ANYCASTの冗長化のデメリット
 - BGPはBest Pathで経路選択されるため、NWの構成を意識してバランシングする必要がある
- IP ANYCASTにて対応出来ない問題について対処する必要がある
 - Health Check Daemonを実装

JANOG48: Refresh DNS Infrastructure with Modern Datacenter Network

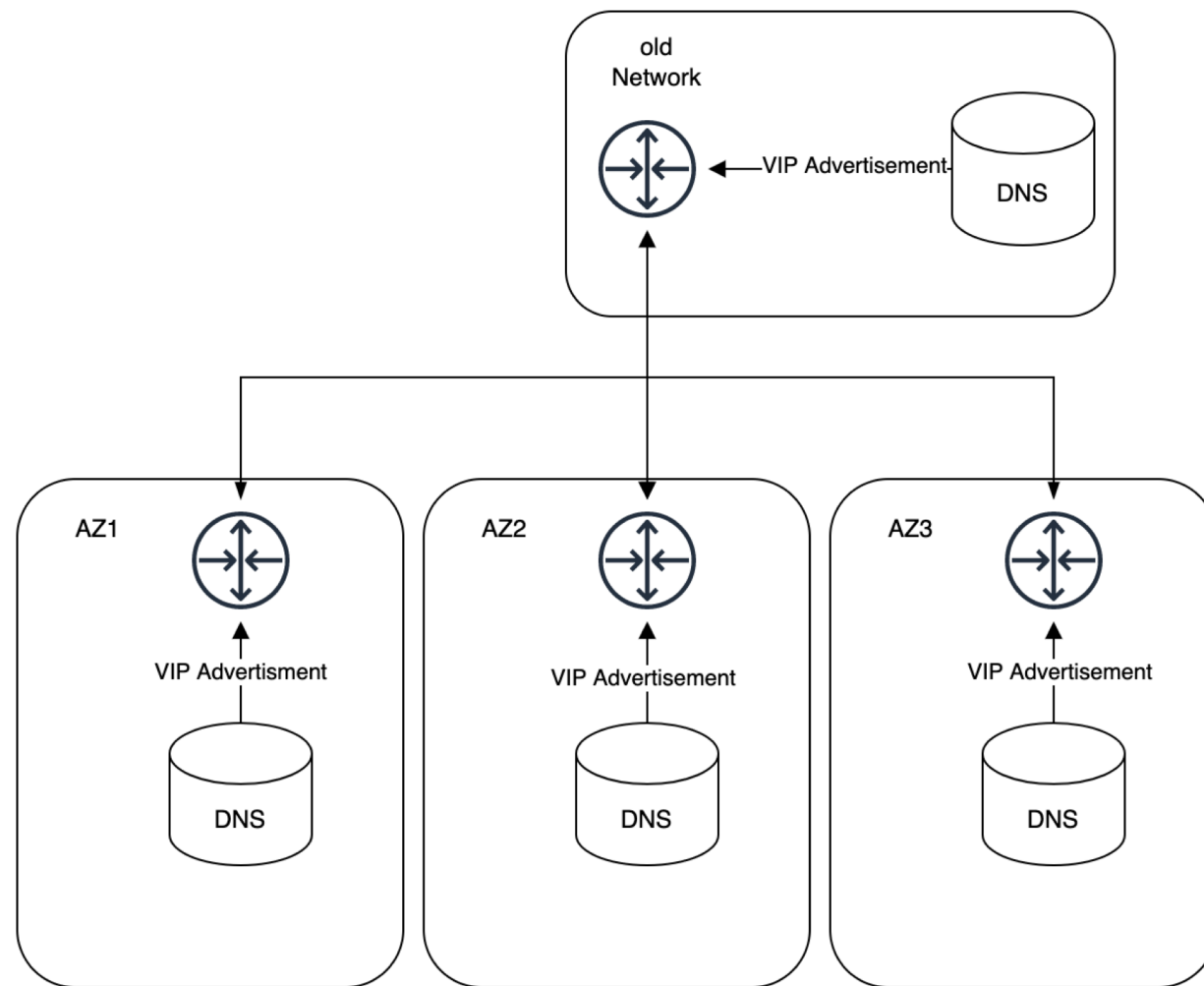
DNS: Multi-AZs Support(1/2)

各AZにDNSサーバをホスティング

- DNSのVMを配置するだけで完了
- AZを指定しVMを作成しデプロイするだけ

AZへのマイグレーションはBGPを用いて

- DNSでは /32 でVIPが広報されている
- DNSのトラフィックは莫大というわけではないので、AZ内での経路広報が安定した段階で切り替え可能



DNS: Multi-AZs Support(2/2)

各AZにDNSサーバをホスティング

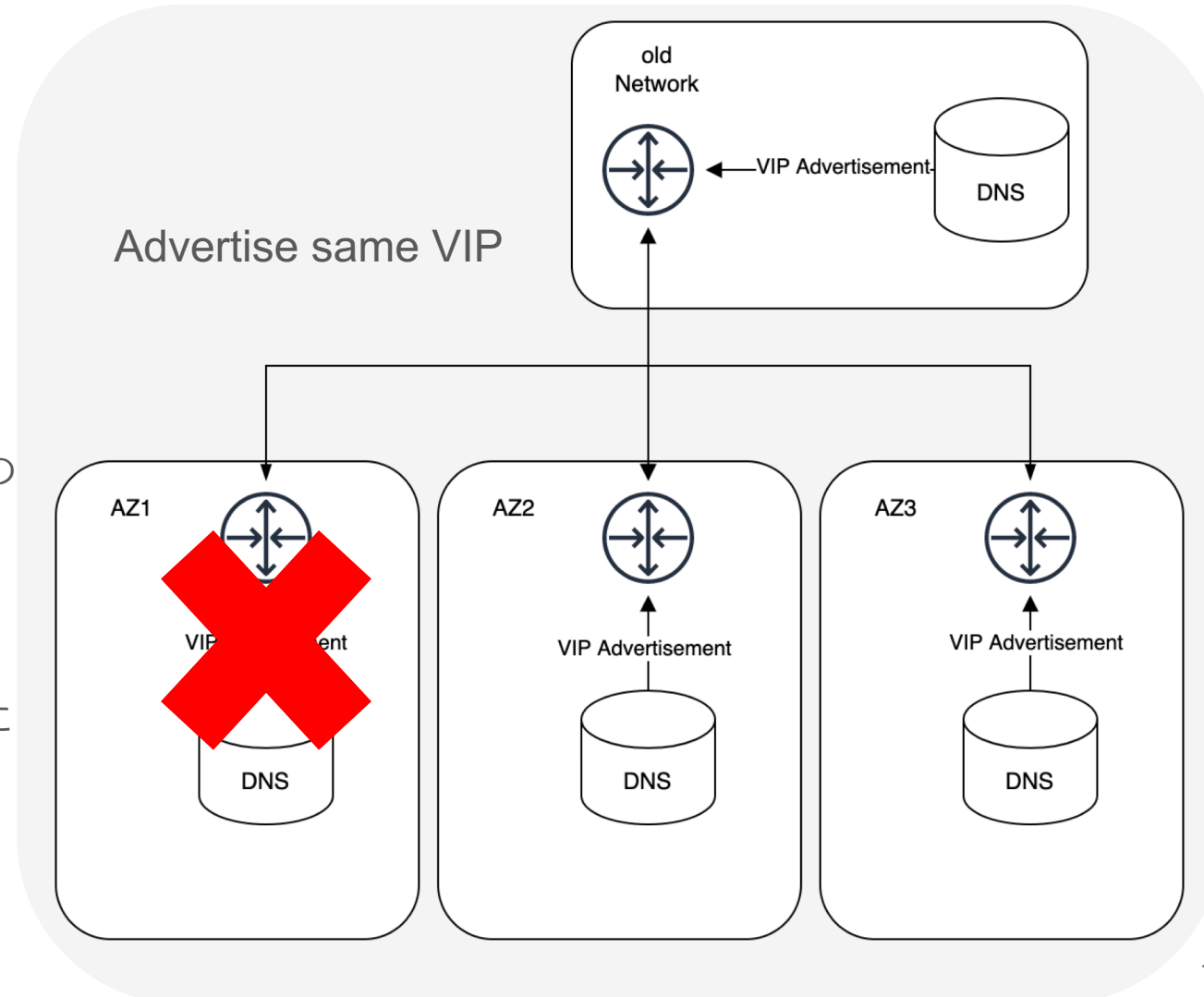
- DNSのVMを配置するだけで完了
- AZを指定しVMを作成しデプロイするだけ

AZへのマイグレーションはBGPを用いて

- DNSでは /32 でVIPが広報されている
- DNSのトラフィックは莫大というわけではないので、AZ内での経路広報が安定した段階で切り替え可能

全てのAZから同じVIPを広報

- AZ Down時には自動的に稼働中のDNS ServerにRoutingが行われる
- DNSサービス自身がDNSの動作をCheckしている



Multi-AZs Support in DNS

- AZ対応を意識して構築されていたため対応が容易
 - Full L3 NetworkでVIPのポータビリティが向上する
- サービスイン時には気を遣う必要がある
 - BGPなので経路の収束などをある程度気を遣う必要がある

Our NFV Services

DNS

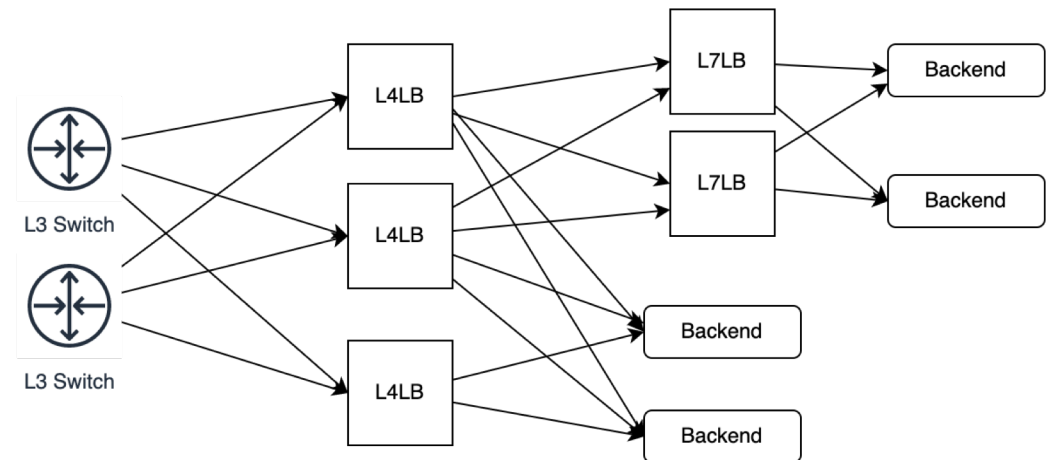
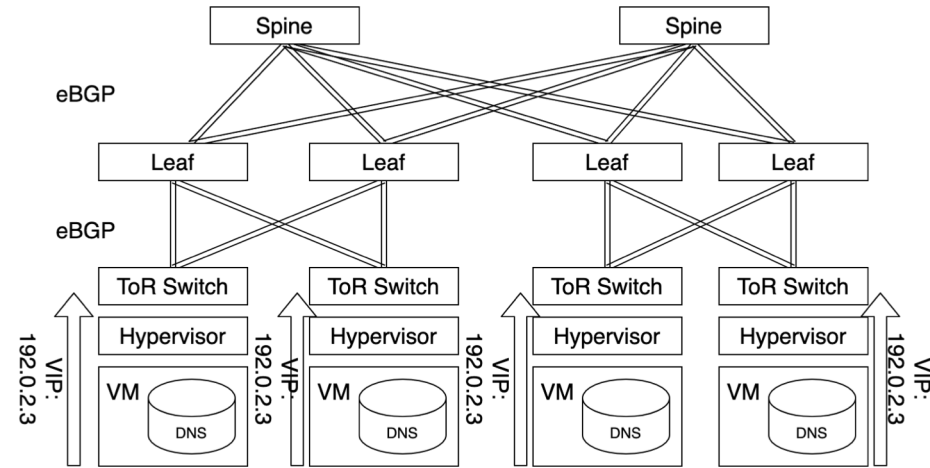
- DNS CacheとDNS Authを提供
- IP ANYCASTを用いて冗長化

Load Balancer

- L4LBとL7LBを提供
- L4LBはIP ANYCASTを用いて冗長化

NAT

- Internetへ向けてのNATサービスを提供
- NATのD-PlaneはIP ANYCASTを用いて分散



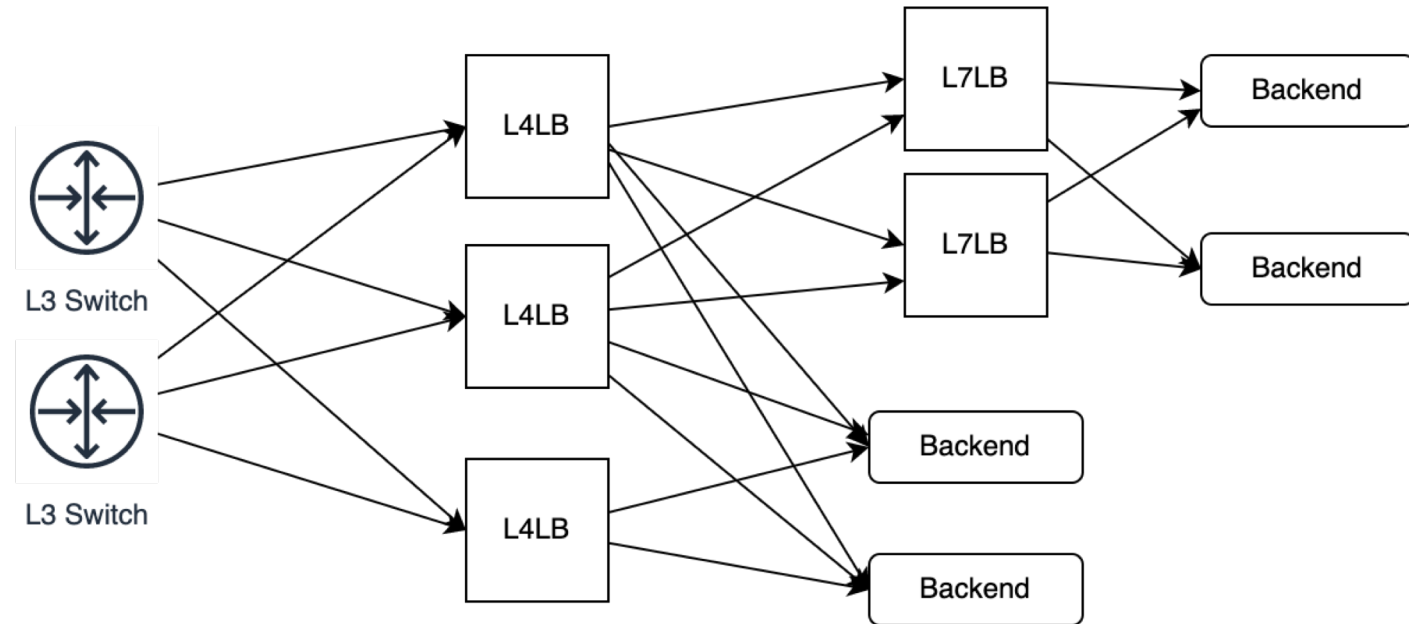
LB: Architecture

L4LB

- 物理サーバ
- LBで管理しているPrefixを広報
- IP ANYCASTを用いて冗長化

L7LB

- k8sのPodsとして構築
- L4LBのバックエンドとして動作する



LB: AZ Support Consideration Points

- **トラフィックバランシング**
 - LBはDC間のトラフィックを発生させる可能性があるNFVサービス
 - ユーザに使い勝手のよいトラフィック分散を検討する
- **サービスイン**
 - LBはDNSと比べてトラフィックが多い
 - サービスインのタイミングでトラフィックがバーストする可能性が高い

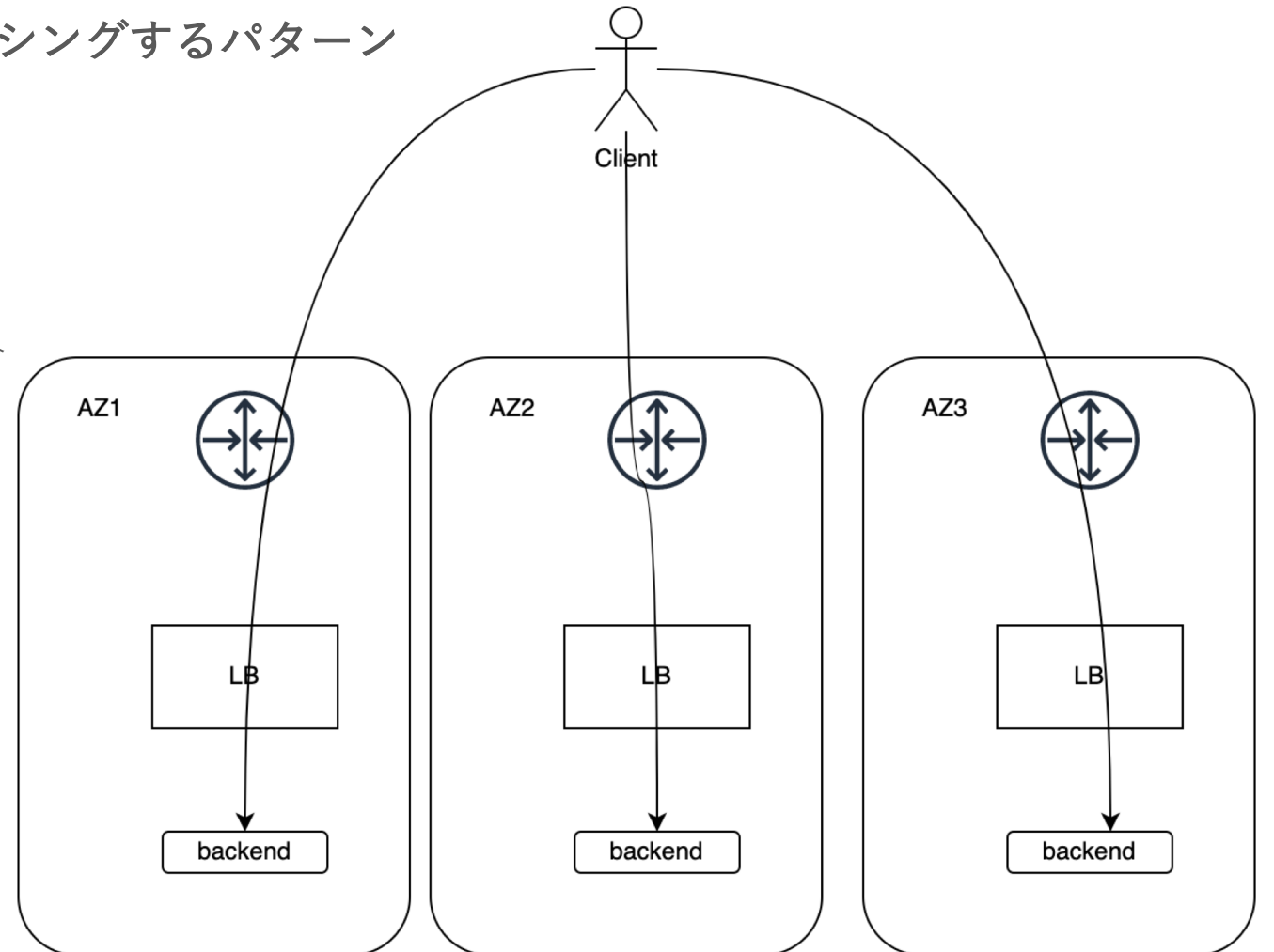
LB: Traffic Balancing(1/2)

AZ内のバックエンドに対してbalancingするパターン

- DC間のトラフィックを削減
- レイテンシが最もすくない状態

問題点

- AZ障害ではない状態でLBがDownした場合に同一AZのBackendが利用されない
- AZ内でのサービスの利用者が多い場合にユーザがCapacityを管理する必要がある
- LBの設定が複雑になる



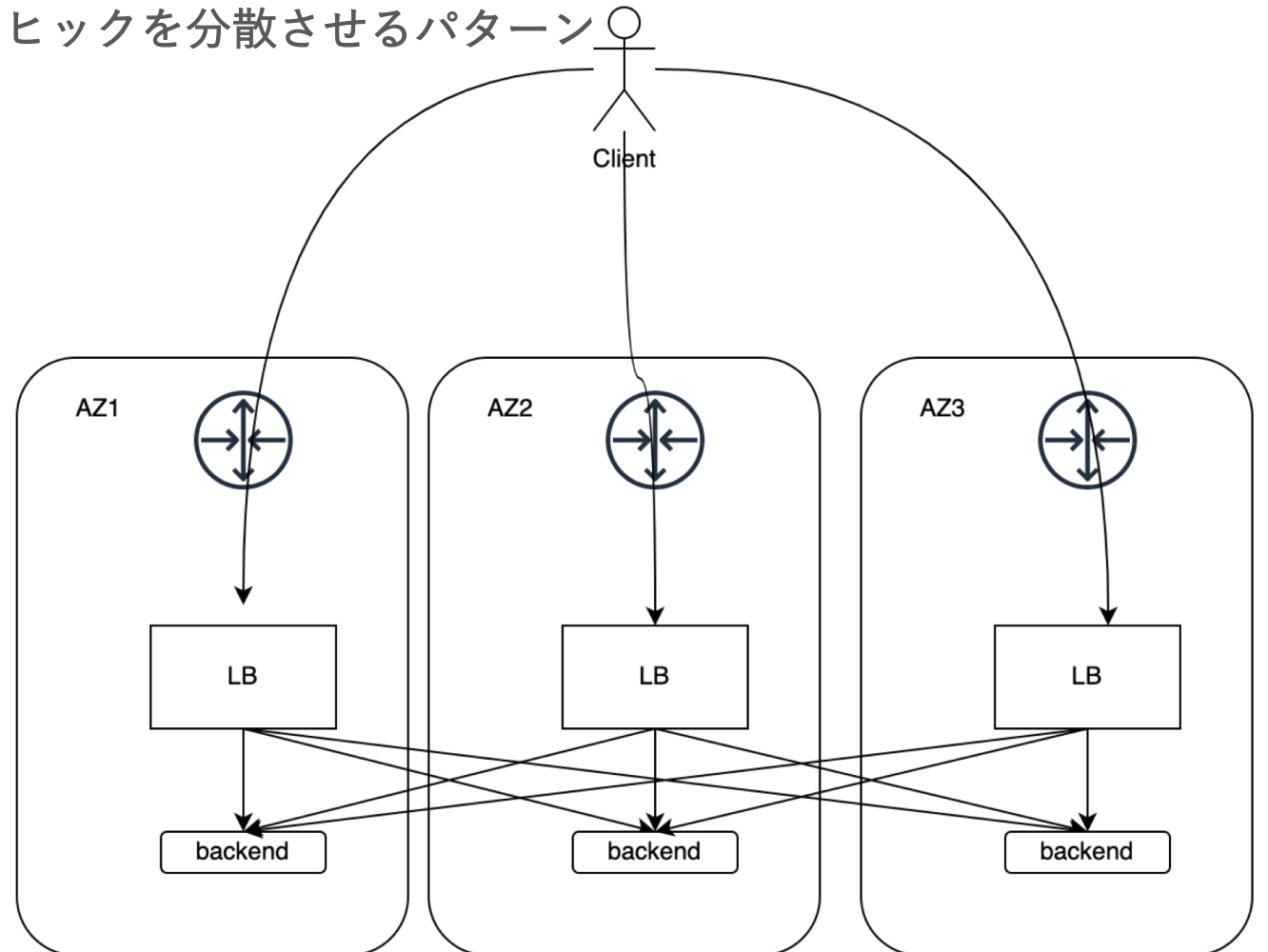
LB: Traffic Balancing(2/2)

AZ関係なく全てのバックエンドにトラフィックを分散させるパターン

- ユーザがAZを意識する必要が無い
- 全てのAZのLBで共通の設定を利用可能

問題点

- DC間のトラフィックが増加
- Latencyが増加する可能性がある



Pros

AZ内で balancing

- DC間のトラフィックを削減する事が出来る
- DC間の通信によるレイテンシの増加がない

Cons

- AZ内でのトラフィックの差をユーザーが意識をする必要がある
- LBのAZ障害時にバランスされないバックエンドが発生する
- LBに複雑なコンフィグレーションが必要

全てのバックエンド にbalancing

- ユーザーがAZでのCapacityを意識する必要が無い
- 全てのAZのLBで共通のコンフィグレーションを利用出来る

- DC間のトラフィックが発生する
- DC間での通信でレイテンシーが増加する可能性がある

Pros

AZ内で balancing

- DC間のトラフィックを削減する事が出来る
- DC間の通信によるレイテンシの増加がない

Cons

- AZ内でのトラフィックの差をユーザーが意識をする必要がある
- LBのAZ障害時にバランスされないバックエンドが発生する
- LBに複雑なコンフィグレーションが必要

全てのバックエンド にbalancing

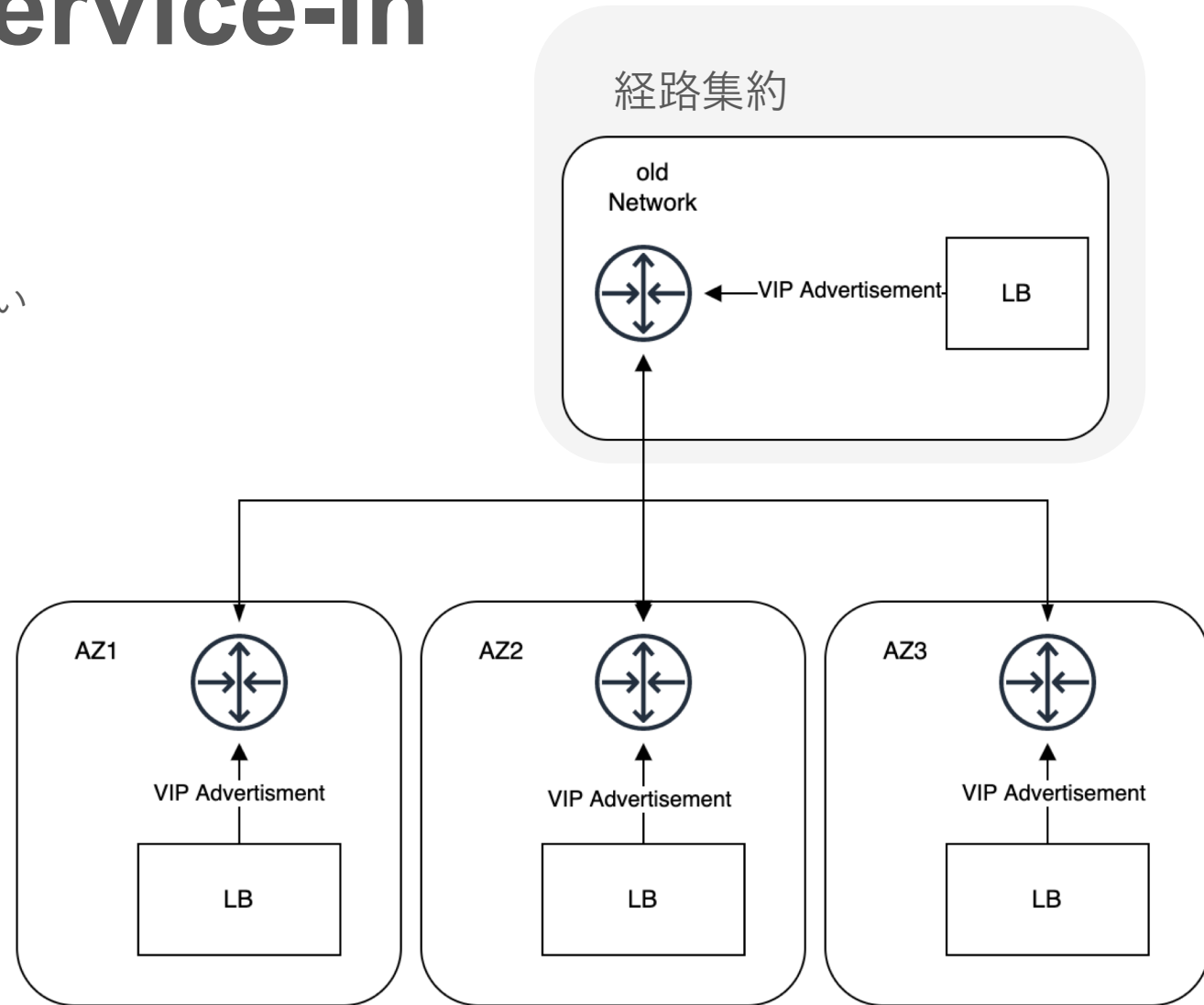
- ユーザーがAZでのCapacityを意識する必要が無い
- 全てのAZのLBで共通のコンフィグレーションを利用出来る

- DC間のトラフィックが発生する
- DC間での通信でレイテンシーが増加する可能性がある

LB: Service-in

トラフィックの曲げ方

- Old NetworkではLBの経路を集約している
- AZ環境ではLBで広報している経路を集約しない事で引き込む



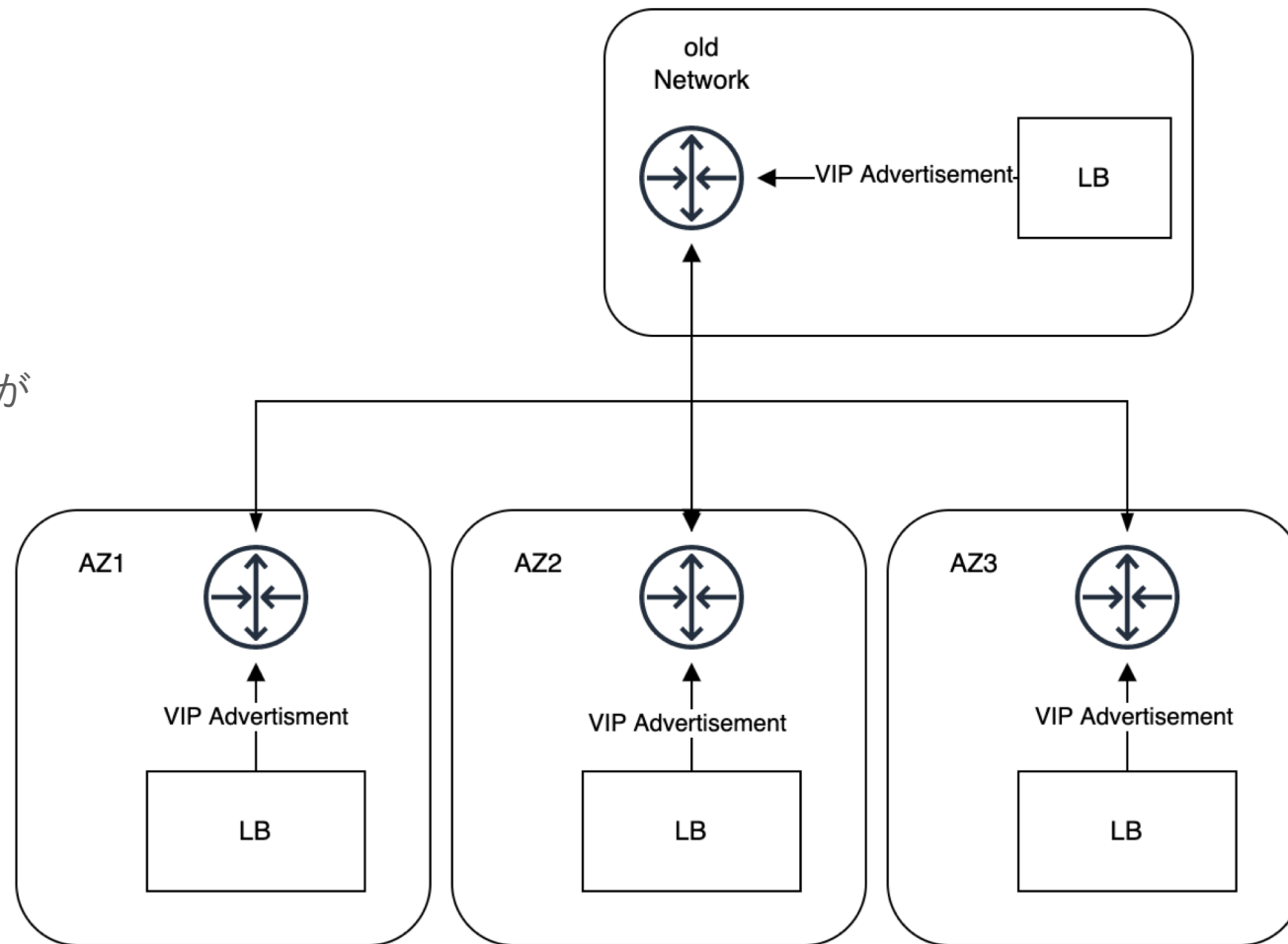
LB: Service-in

トラフィックの曲げ方

- Old NetworkではLBの経路を集約している
- AZ環境ではLBで広報している/27を集約しない事で引き込む

トラフィックのバースト

- 経路広報の開始時に1台のホストにトラフィックが集中し、障害が発生する可能性がある
- 1つのLBクラスターで90Gbpsを記録



LB: Service-in

トラフィックの曲げ方

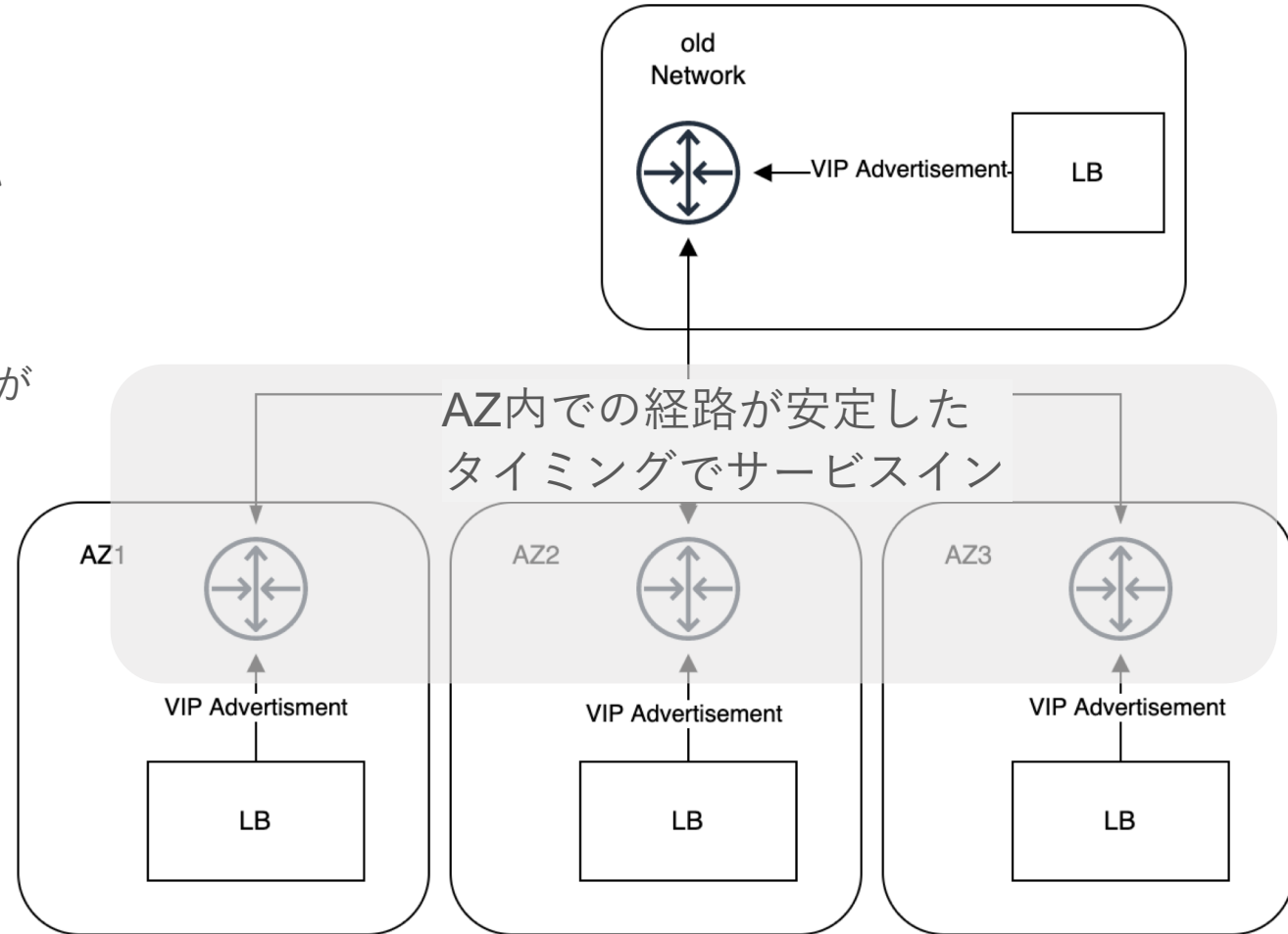
- Old NetworkではLBの経路を集約している
- AZ環境ではLBで広報している/27を集約しない事で引き込む

トラフィックのバースト

- 経路広報の開始時に1台のホストにトラフィックが集中し、障害が発生する可能性がある
- 1つのLBクラスターで90Gbpsを記録

解決策

- NWチームとオペレーションを行う事で解決
- AZ内でのみトラフィックが閉じるように設計し、安全にサービスイン出来る状態にする



Conclusion

- VerdaのNFVサービスはN+1構成で提供
 - DNS
 - BGPでIP ANYCAST
 - LB
 - BGP + L4とL7の多段構成
- スケール出来る構成はマイグレーションも容易な構成
- ユーザが気にする事なくAZを提供する事で、LINEのサービス全体の可用性を上げる
- LINEのインフラ全体で協力して安定したサービスを提供