

VPP Accelerated Cloud Native Load Balancer

2018/10/23

立見祐介 (Tatsumi, Yusuke) @ Yahoo Japan Corporation



**Please find the Japanese slide
after English version.**

Yahoo! JAPAN introduction

The screenshot shows the main interface of the Yahoo! JAPAN website. At the top, there's a navigation bar with links for 'Yahoo! JAPAN', 'トップ', 'アプリ', 'トラベル', 'サフオク!', 'ショッピング', 'YAHOO! JAPAN' logo, 'Premium', 'カード', and 'メール'. Below the logo is a search bar with the placeholder 'Q 検索'. The main content area features several sections: 'Yahoo! JAPAN アプリ タブレット版 iPad / Android' with a download button; '明日の雨をお知らせ!' (Rain forecast); 'ログイン' (Login) with options for 'IDでもっと便利に' (More convenient with ID), 'メールアドレス取得', 'ポイント確認', and 'カード新規登録'; '日 2018年9月7日 (金)' (Date: Friday, September 7, 2018); '今日の天気 (東京)' (Weather in Tokyo) showing 10% chance of rain at 19°C/16°C; '明日の天気 (東京)' (Weather in Tokyo) showing 10% chance of rain at 21°C/16°C; '運行情報' (Travel information) with a note '事故・遮断情報はありません。(18時06分)' (No accidents or roadblocks at 18:06); 'スコアボード' (Scoreboard) with tabs for 'プロ野球' (Professional baseball) and 'Jリーグ' (J League); and various other news and service links like '主なサービス', '動画', 'ニュース', '経済', 'エンタメ', 'スポーツ', 'カレンダー', '天気', '災害', 'スポーツナビ', 'ファイナンス', 'テレビ', 'GIGA', 'ゲーム', 'Yahoo!モバゲー', '地図', '路線情報', 'しごと検索', '不動産', '自動車', 'testman', 'ブログ', 'TRILL', 'パートナー', 'お気に入り', 'ブックストア', '映画', '点い', 'ロコ', 'ネットで話題の無料動画', and 'ガラスに映る自分を見た猫の優しい行動'.

The screenshot shows the mobile application interface for iOS. At the top, it displays the time '7:25' and battery level '100%'. The search bar contains 'Yahoo! 検索' with a microphone icon and a blue '検索' (Search) button. Below the search bar is a grid of icons: 'メール' (Email) with a red notification '2', '24/18 100% 港区' (Port City 100%), '魚座 82点' (Pisces 82 points), 'お気に入り' (Favorites), 'ショッピング' (Shopping); 'スポーツナビ' (Sports Navigation), 'ヤフオク!' (Yahoo Auctions), '路線' (Route), '旅行・出張' (Travel/Outing); and tabs for 'フォロー' (Follow), 'すべて' (All), 'ニュース' (News), '話題' (Topics), '芸能' (Entertainment), and 'スポーツ' (Sports). The main content area lists news articles with images and publication dates: '敷金やツケ 民法どう変わる?' (New rules for deposits and fees? - NEW 9/28(月) 7:23), '野良猫に餌やり禁止 賛否続々' (Ban on feeding feral cats - NEW 9/28(月) 6:55), 'トルコ最大級の地下都市発見' (Discovery of the largest underground city in Turkey - 9/28(月) 6:31), and '150年ぶり サボテン新種発見' (Discovery of a new cactus species after 150 years - NEW 9/28(月) 7:08).

Photo coordinate by AFLO

Yahoo! JAPAN introduction



Photo coordinate by AFLO

https://s.yimg.jp/i/docs/ir/archives/jp/nenji/2017/jp2018integrated_report.pdf

Background and issues of LB

Demands for speed

- Baremetal server: VIP management by operation
- VM: LBaaS (automation system) manages VIPs
- Container: much faster VIP update

Explosive demands

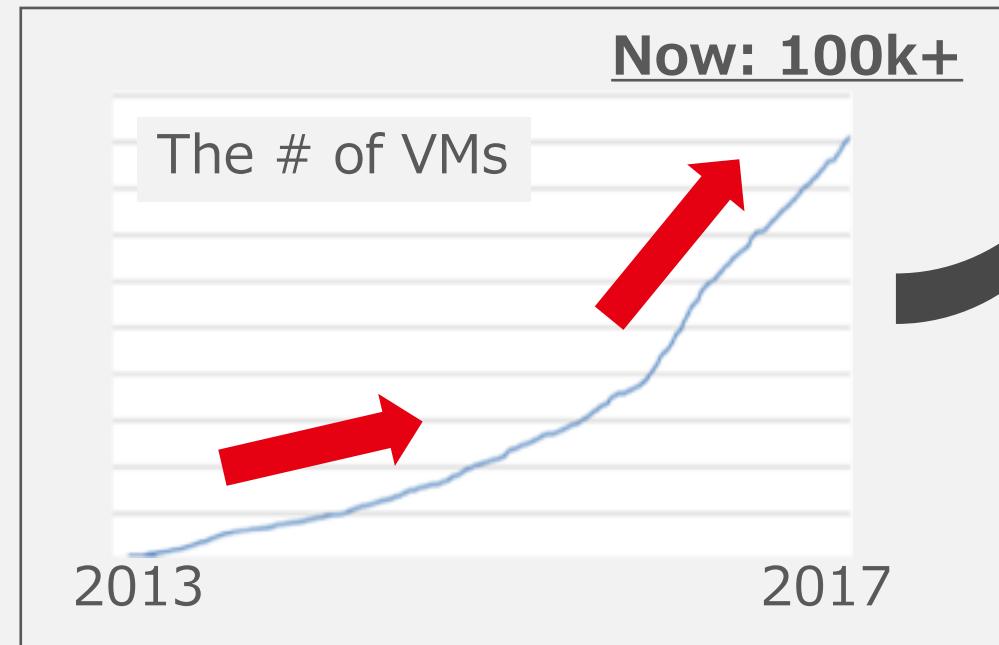
- News, Weather
- Shopping
- Settlement
- Etc, etc,

Background and issues of LB

Demands for speed

- Baremetal server: VIP management by operation
- VM: LBaaS (automation system) manages VIPs
- Container: much faster VIP update

Explosive demands
• News, Weather
• Shopping
• Settlement
• Etc, etc,



Background and issues of LB

Total responsibility of Load Balancer =

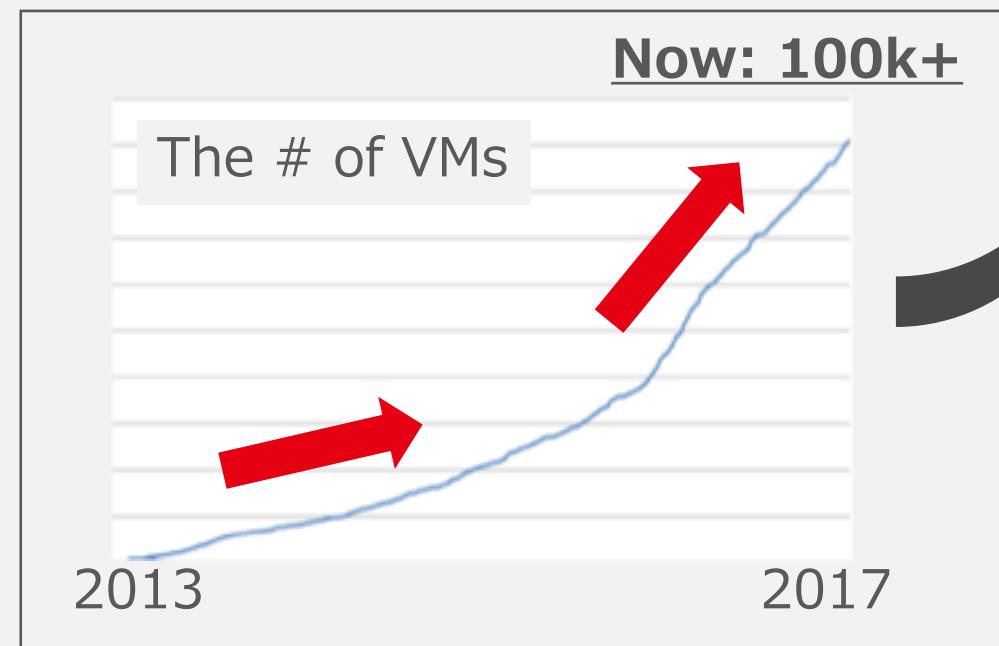
Demands for speed

- Baremetal server: VIP management by operation
- VM: LBaaS (automation system) manages VIPs
- Container: much faster VIP update



Explosive demands

- News, Weather
- Shopping
- Settlement
- Etc, etc,



Background and issues of LB

Total responsibility of Load Balancer =

Demands for speed

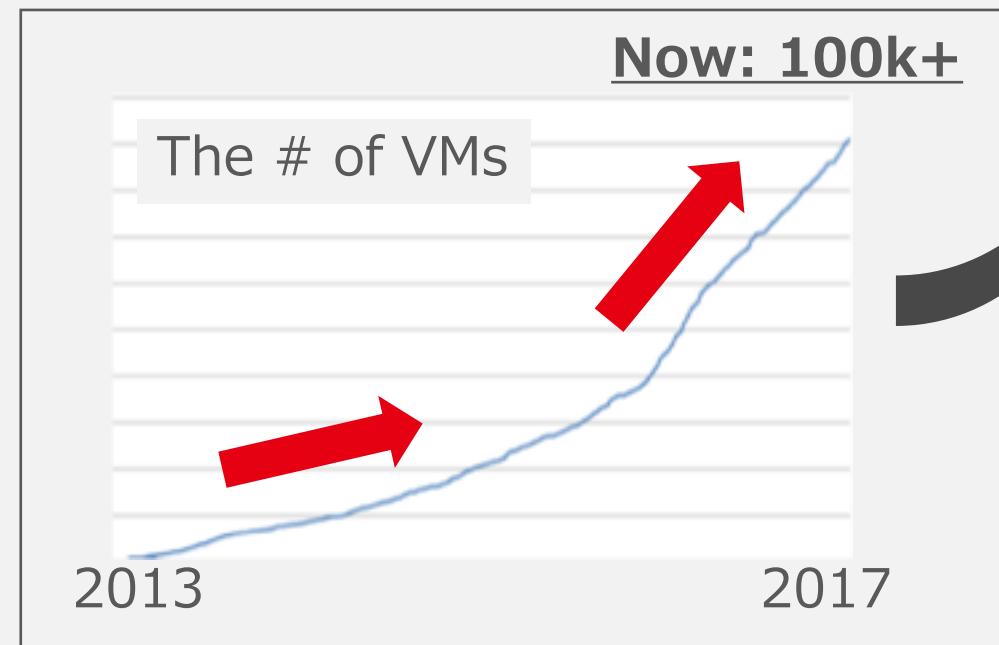
- Baremetal server: VIP management by operation
- VM: LBaaS (automation system) manages VIPs
- Container: much faster VIP update



Explosive demands

- News, Weather
- Shopping
- Settlement
- Etc, etc,

**Is LBaaS going well?
=> Issue still exists**



Limitation in our use case, and OSS

- Our appliance based LBaaS

Pros

- High performance
- Vender supported
- Verified features

Cons

- Lead time
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- Layer 2 limitation (not on CLOS)

Limitation in our use case, and OSS

- Our appliance based LBaaS

Pros

- High performance
- Vender supported
- Verified features

Cons

- Lead time
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- Layer 2 limitation (not on CLOS)

- OSS based LBaaS

Pros

- Commodity server/NIC
- Traceability/Debuggability
- Easy to add features
- Integration with other OSS

Cons

- No dedicated support
- Verify your self
- Slow packet forwarding (?)

Limitation in our use case, and OSS

- Our appliance based LBaaS

Pros

- High performance
- Vender supported
- Verified features

Cons

- Lead time
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- Layer 2 limitation (not on CLOS)



- OSS based LBaaS

Pros

- Commodity server/NIC
- Traceability/Debuggability
- Easy to add features
- Integration with other OSS

Cons

- No dedicated support
- Verify your self
- Slow packet forwarding (?)

Collaboration with VPP breaks our limitations in our use case!

Limitation in our use case, and OSS

- Our appliance based LBaaS

Pros

- High performance
- Vender supported
- Verified features

Cons

- Lead time
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- Layer 2 limitation (not on CLOS)



- OSS based LBaaS

Pros

- Commodity server/NIC
- Traceability/Debuggability
- Easy to add features
- Integration with other OSS

Cons

- No dedicated support
- Verify your self
- Slow packet forwarding (?)

Collaboration with VPP breaks our limitations in our use case!

?

VPP overview

VPP is a OSS “switch” by FD.io community

- Vector Packet Processing
 - Optimal CPU utilization by Vectorizing packets
- DPDK
 - Fast/stable packet forwarding by kernel bypass
- Plugin extension
 - Switching, Routing, ACL, NAT, SRv6, LB, etc
- Integration
 - ODL, OPNFV, OpenStack, k8s, etc
- CSIT: func/perf CI in community



VPP overview

VPP is a OSS “switch” by FD.io community

- Vector Packet Processing
 - Optimal CPU utilization by Vectorizing packets
- DPDK
 - Fast/stable packet forwarding by kernel bypass
- Plugin extension
 - Switching, Routing, ACL, NAT, SRv6, LB, etc
- Integration
 - ODL, OPNFV, OpenStack, k8s, etc
- CSIT: func/perf CI in community



- 13:30-13:55 Accelerate Cloud Native with FD.io *Naoyuki Mori & Ping Yu, Intel*
- Running VPP machine at Intel booth.

Our L4 Load Balancer requirements and implementation

Requirements

- OSS based and fast data plane
- L3DSR LB (using IPv4 DSCP field)
- Scaling-in/out capability
- Integrated with ECMP/BGP on CLOS

Our L4 Load Balancer requirements and implementation

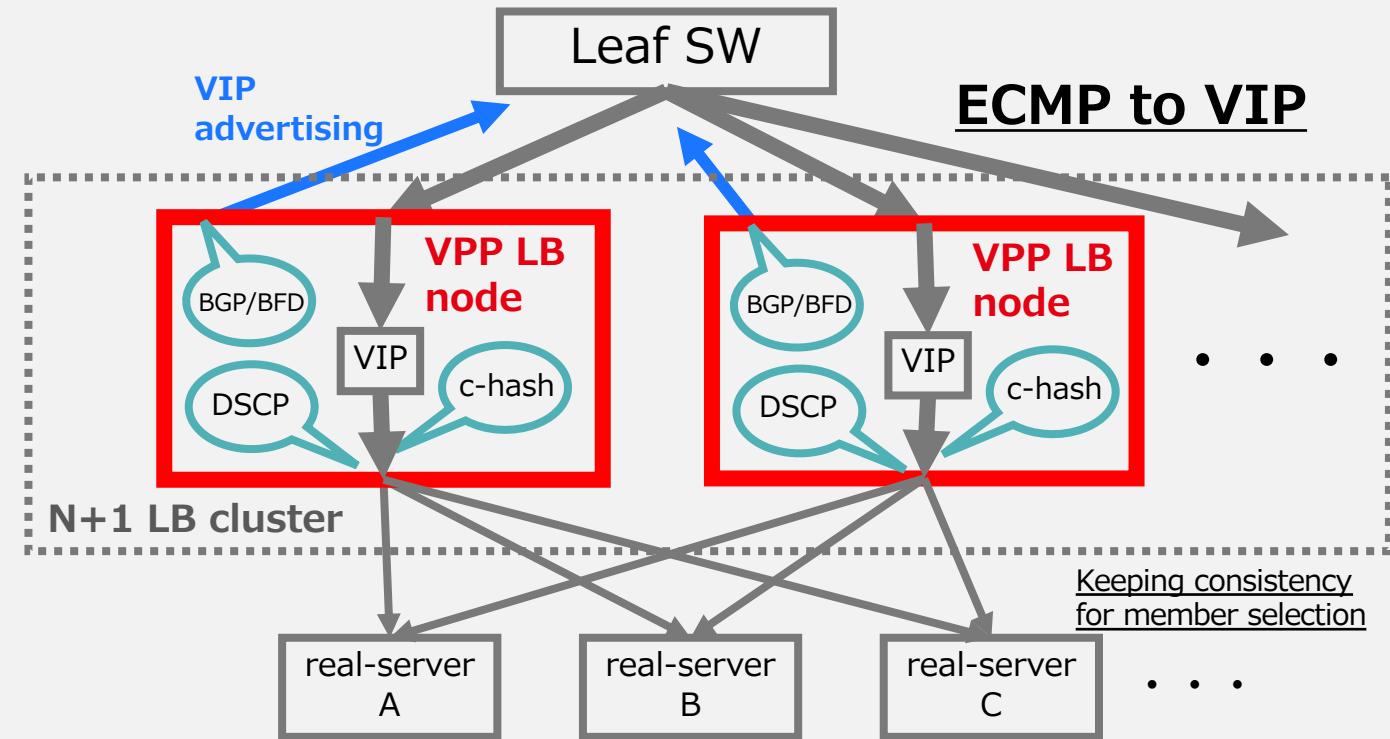
Requirements / Implementation

- OSS based and fast data plane
 - VPP
- L3DSR LB (using IPv4 DSCP field)
 - VPP LB plugins
- Scaling-in/out capability
 - Consistent-hash feature
- Integrated with ECMP/BGP on CLOS
 - Router plugin on CLOS network

Our L4 Load Balancer requirements and implementation

Requirements / Implementation

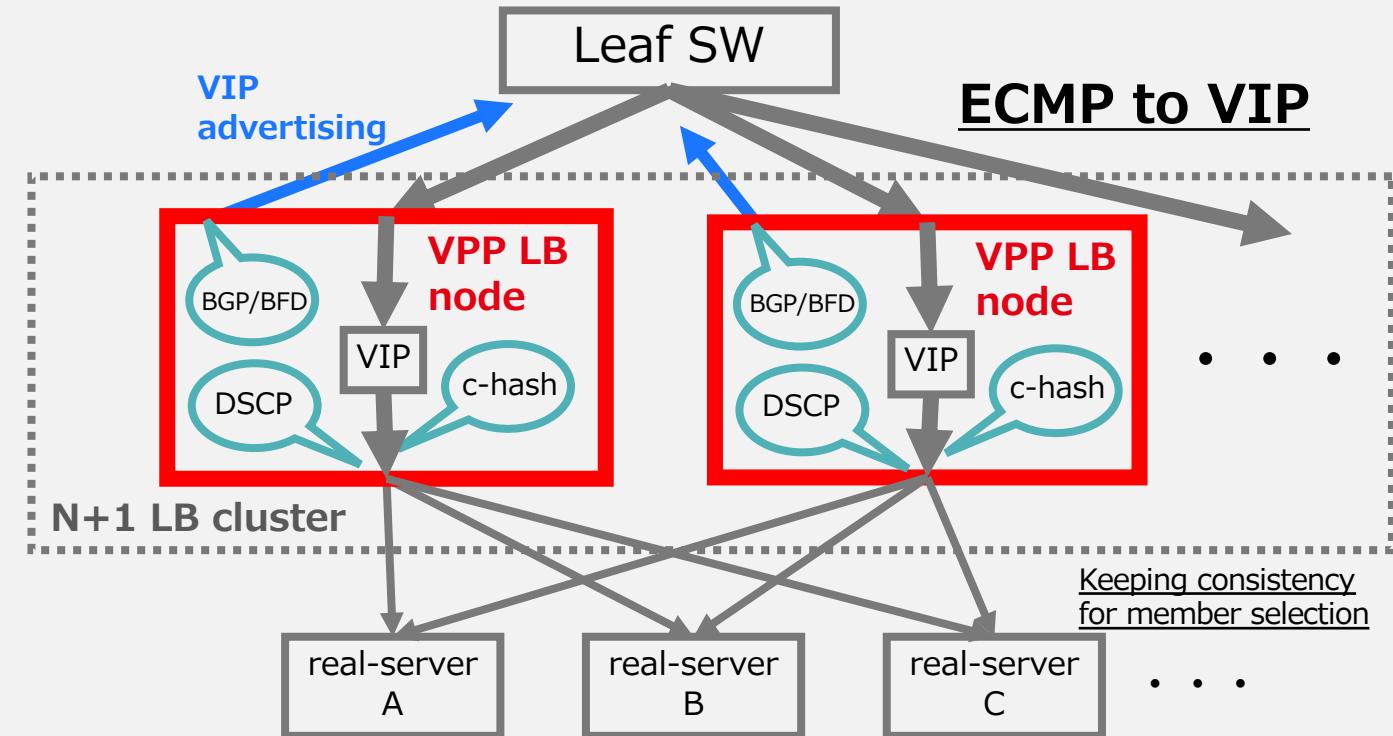
- OSS based and fast data plane
 - VPP
- L3DSR LB (using IPv4 DSCP field)
 - VPP LB plugins
- Scaling-in/out capability
 - Consistent-hash feature
- Integrated with ECMP/BGP on CLOS
 - Router plugin on CLOS network



Our L4 Load Balancer requirements and implementation

Requirements / Implementation

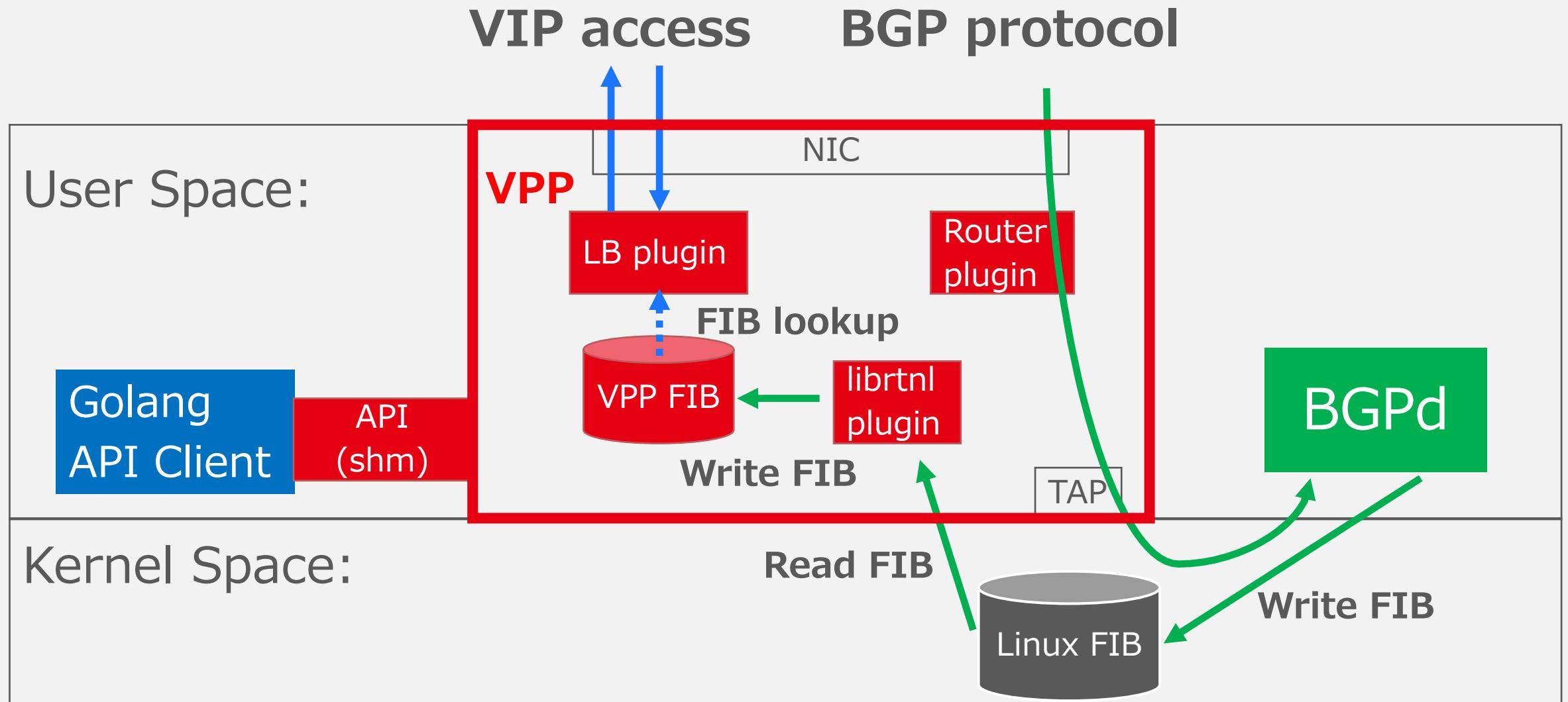
- OSS based and fast data plane
 - VPP
- L3DSR LB (using IPv4 DSCP field)
 - VPP LB plugins
- Scaling-in/out capability
 - Consistent-hash feature
- Integrated with ECMP/BGP on CLOS
 - Router plugin on CLOS network



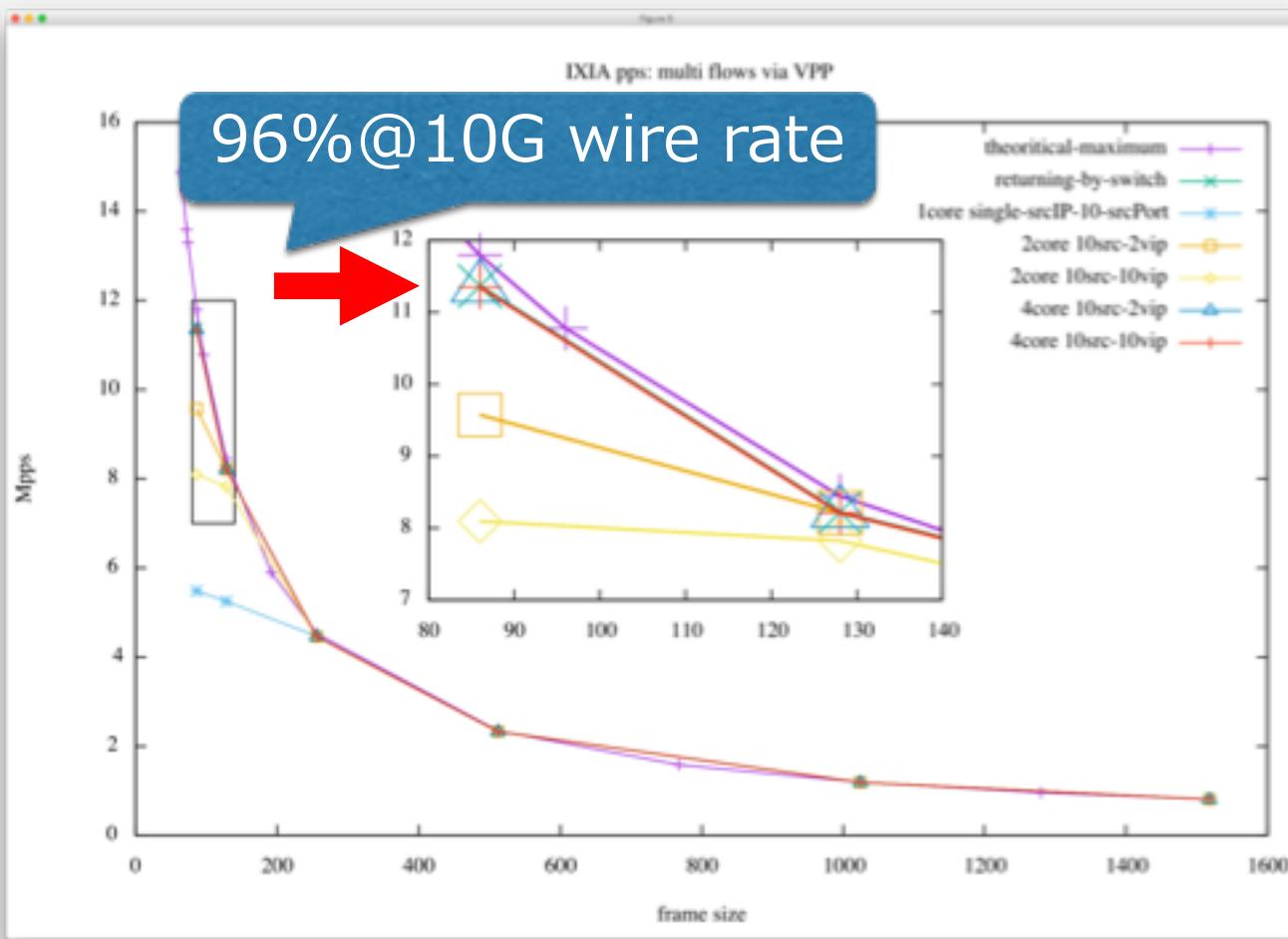
VPP Realized requirements of our use case:

- OSS based stateless L3DSR L4 Load Balancer
- Stable scaling-in/out N+1 capability
- Layer2 agnostic, easy EoL (life cycle) management

Internal VPPLB architecture overview

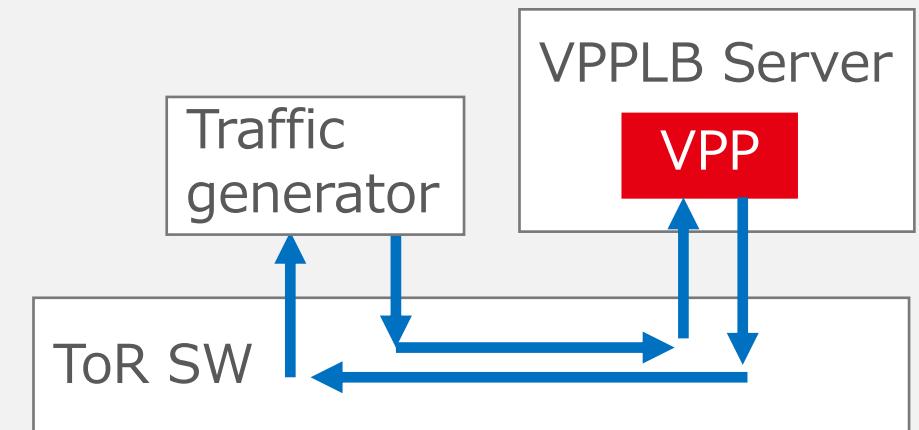


VPPLB/node throughput test: 10G



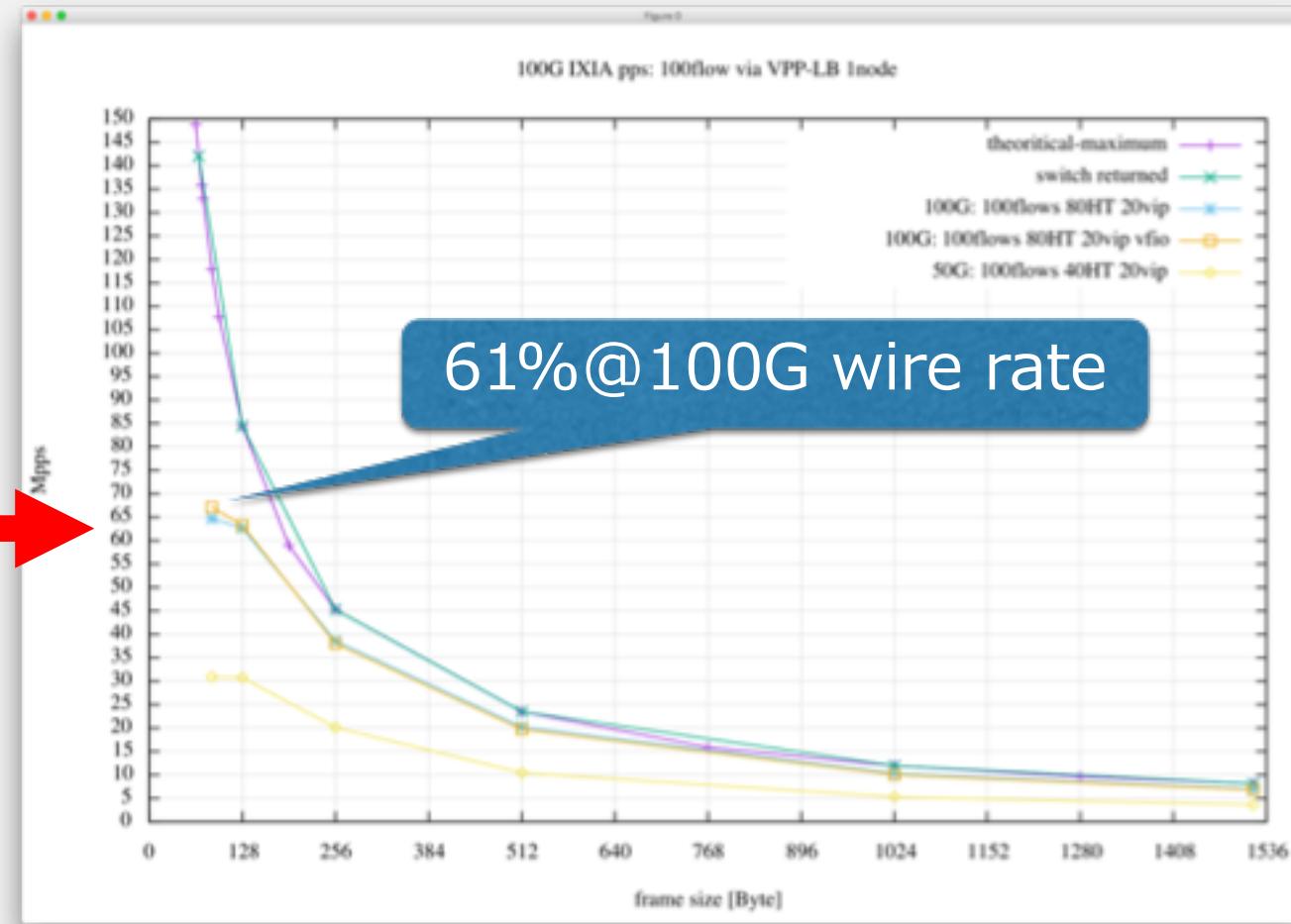
Traffic generator: Ixia
VPP Server Spec:

- CPU: Xeon E5-2650L v3 * 2
- Memory: 384GB
- NIC: Intel X540-AT2



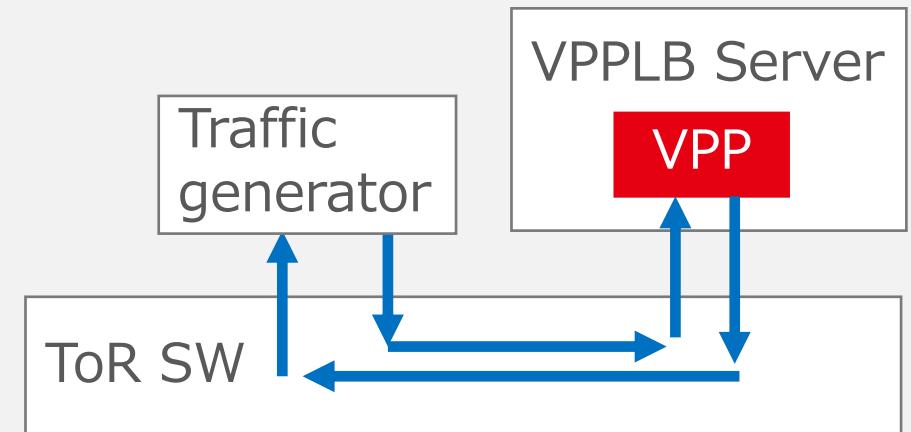
Sufficient performance on 10Gbps server!

VPPLB throughput test: 100G



Traffic generator: Ixia
VPP Server Spec:

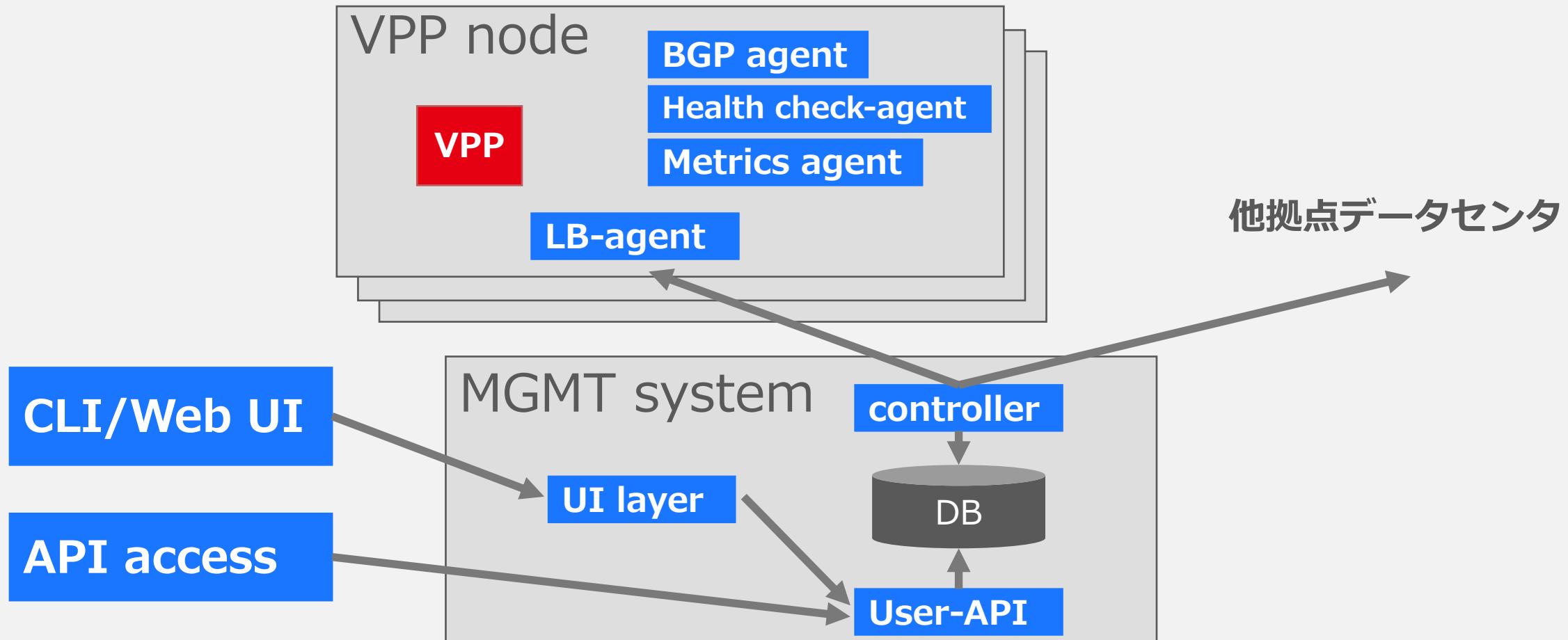
- CPU: Xeon Gold 6152 * 2
- Memory: 754GB
- NIC: Intel XXV710-DA2 * 2



Challenging to the next mile stone!

LB management System

- In-house integrated management system



Realizing fast and stable VIP deployment

Call to Action

- FD.io: A Good open community resolving real-world networking issues, and updating features
 - Quick response from community
 - Deep networking knowledge in community
- Collaboration
 - Sharing your ideas/issues
 - Coding and verifying in each environment
 - Talk to each other

Let's work together!

**Thank you
for your listening!**



VPPで実現する Cloud Nativeロードバランサ

2018/10/23

立見祐介 (Tatsumi, Yusuke) @ ヤフー株式会社



Yahoo! JAPAN introduction

The screenshot shows the Yahoo! JAPAN desktop homepage. At the top, there's a navigation bar with links for 'Yahoo! JAPAN' (with a lock icon), 'トップ', 'アプリ版', 'トライブ', 'サフォク!', 'ショッピング', 'YAHOO! JAPAN' logo, 'ヘルプ', 'Premium', 'カード', and 'メール'. Below the navigation is a search bar with placeholder text 'Q 検索' and a blue '検索' button. The main content area features several sections: 'Yahoo! JAPAN アプリ タブレット版 iPad / Android 無料' (with a screenshot showing the app interface); '明日の雨をお知らせ!' (Weather forecast for tomorrow); 'ログイン' (Login) with options for 'IDでもっと便利に' (New features), 'メールアドレス取得', 'ポイント確認', and 'カード新規登録'; '毎日更新' (Daily updates) with news items like '敷金やツケ 民法どう変わる?' and '野良猫に餌やり禁止 賛否続々'; 'あなたにおすすめ' (Recommended for you) featuring products like '思ひ出しきくなっちゃう、おもしろ文具' and 'おしゃれなティーポットで優雅なティータイム'; and 'ネットで話題の無料動画' (Free video topics) with a thumbnail for 'ガラスに映る自分を見た猫の優しい行動'.

The screenshot shows the Yahoo! JAPAN mobile app interface on an iPhone. The top status bar indicates signal strength, battery level (100%), and time (7:25). The main screen has a search bar with 'Yahoo! 検索' and a microphone icon, followed by a blue '検索' button. Below the search bar is a grid of icons: 'メール' (Email) with a red notification badge '2', '24/18 100% 港区' (Weather forecast for Tokyo Bay Area), '魚座 82点' (Horoscope for Pisces), 'お気に入り' (Favorites), and 'ショッピング' (Shopping). Further down are more icons: 'スポーツナビ' (Sports Navigation), 'ヤフオク!' (Yahoo Auctions), '路線' (Route), '旅行・出張' (Travel/Outing), and 'すべて' (All). A navigation bar at the bottom includes tabs for 'フォロー' (Follow), 'すべて' (All), 'ニュース' (News), '話題' (Topics), '芸能' (Entertainment), and 'スポーツ' (Sports).

The main content area displays news articles with images and timestamps:

- 敷金やツケ 民法どう変わる?** (NEW 9/28(月) 7:23)
- 野良猫に餌やり禁止 賛否続々** (NEW 9/28(月) 6:55)
- トルコ最大級の地下都市発見** (9/28(月) 6:31)
- 150年ぶり サボテン新種発見** (NEW 9/28(月) 7:08)

Photo coordinate by AFLO

Yahoo! JAPAN introduction



Photo coordinate by AFLO

ロードバランサ振り返り

スピード要求の変遷

- 物理サーバ時代：VIPを手で運用
- VM時代: LBaaS (自動化システム) でVIP払い出し
- コンテナ時代: より早いVIP払い出しと更新

爆発的なサービス要求

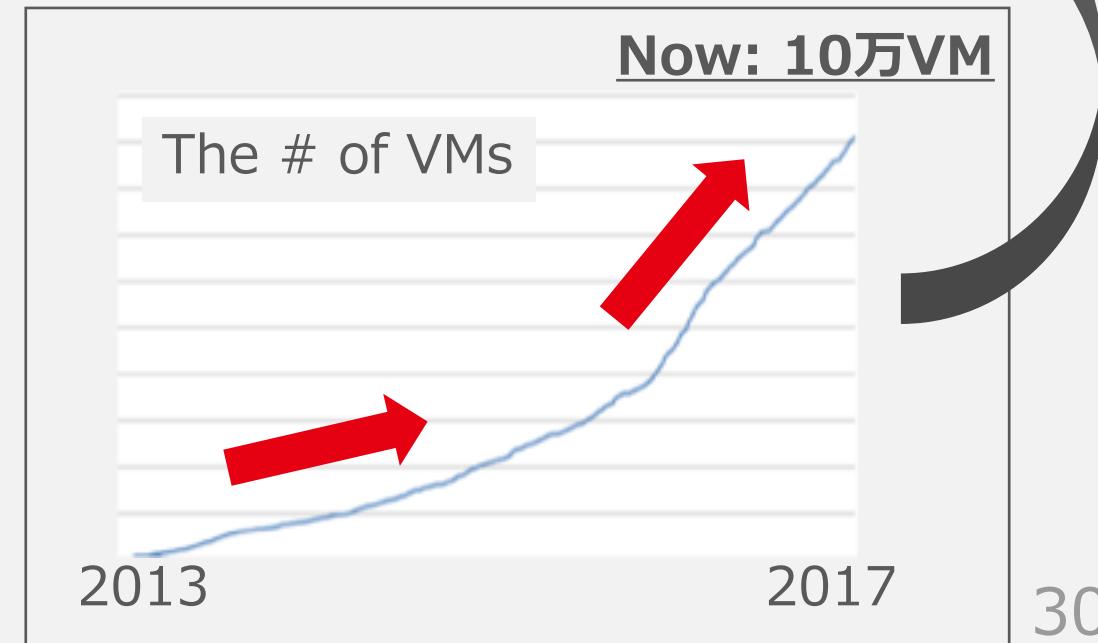
- ニュース、天気
- ショッピング
- 決済
- Etc, etc,

ロードバランサ振り返り

スピード要求の変遷

- 物理サーバ時代：VIPを手で運用
- VM時代: LBaaS（自動化システム）でVIP払い出し
- コンテナ時代: より早いVIP払い出しと更新

- 爆発的なサービス要求
- ニュース、天気
 - ショッピング
 - 決済
 - Etc, etc,



ロードバランサ振り返り

ロードバランサに求められるもの =

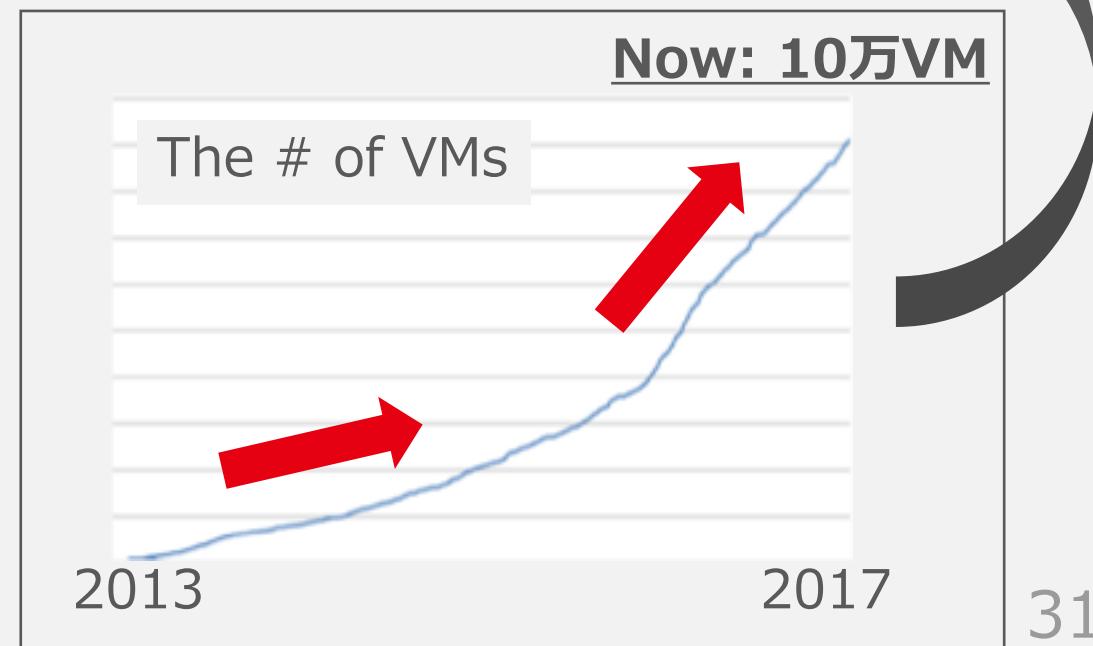
スピード要求の変遷

- 物理サーバ時代 : VIPを手で運用
- VM時代: LBaaS (自動化システム) でVIP払い出し
- コンテナ時代: より早いVIP払い出しと更新



爆発的なサービス要求

- ニュース、天気
- ショッピング
- 決済
- Etc, etc,



ロードバランサ振り返り

ロードバランサに求められるもの =

スピード要求の変遷

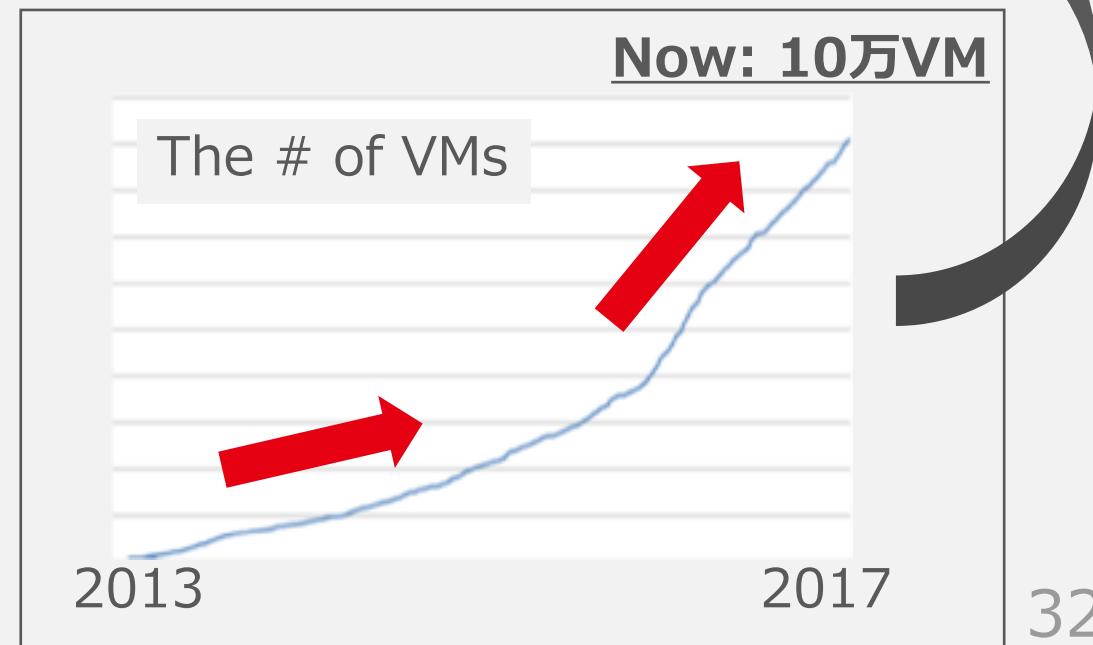
- 物理サーバ時代 : VIPを手で運用
- VM時代: LBaaS (自動化システム) でVIP払い出し
- コンテナ時代: より早いVIP払い出しと更新



爆発的なサービス要求

- ニュース、天気
- ショッピング
- 決済
- Etc, etc,

既存のLBaaSで十分か?
=> 課題が残っている



ヤフーのLBaaSの課題とOSS

- ヤフーのアプライアンスLBaaS

メリット

- High performance
- ベンダーサポート
- 動作確認済み

デメリット

- 納期
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- レイヤー2構成の限界 (not on CLOS)

ヤフーのLBaaSの課題とOSS

- ヤフーのアプライアンスLBaaS

メリット

- High performance
- ベンダーサポート
- 動作確認済み

デメリット

- 納期
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- レイヤー2構成の限界 (not on CLOS)

- OSSを利用したLBaaS

メリット

- Commodity server/NIC
- Traceability/Debuggability
- Easy to add features
- Integration with other OSS

デメリット

- 専属サポート無し
- 自分たちで動作確認
- パケット処理が遅い(?)

ヤフーのLBaaSの課題とOSS

- ヤフーのアプライアンスLBaaS

メリット

- High performance
- ベンダーサポート
- 動作確認済み

デメリット

- 納期
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- レイヤー2構成の限界 (not on CLOS)



- OSSを利用したLBaaS

メリット

- Commodity server/NIC
- Traceability/Debuggability
- Easy to add features
- Integration with other OSS

デメリット

- 専属サポート無し
- 自分たちで動作確認
- パケット処理が遅い(?)

ヤフーのLBaaSの既存制約を乗り越えるVPPと協力

ヤフーのLBaaSの課題とOSS

- ヤフーのアプライアンスLBaaS

メリット

- High performance
- ベンダーサポート
- 動作確認済み

デメリット

- 納期
- Hardware EoL
- Black box
- Scaling-up, no so scaling-out
- レイヤー2構成の限界 (not on CLOS)



- OSSを利用したLBaaS

メリット

- Commodity server/NIC
- Traceability/Debuggability
- Easy to add features
- Integration with other OSS

デメリット

- 専属サポート無し
- 自分たちで動作確認
- パケット処理が遅い(?)

ヤフーのLBaaSの既存制約を乗り越えるVPPと協力
?

VPP overview

VPP is a OSS “switch” by FD.io community

- Vector Packet Processing
 - ベクトル化されたパケット処理でCPUを効率的に利用
- DPDK
 - 高速・安定なパケットフォワーディング
- Pluginでの機能拡張
 - Switching, Routing, ACL, NAT, SRv6, LB, etc
- 各種連携
 - ODL, OPNFV, OpenStack, k8s, etc
- CSIT: 機能・性能をコミュニティ内のCIで担保



VPP overview

VPP is a OSS “switch” by FD.io community

- Vector Packet Processing
 - ベクトル化されたパケット処理でCPUを効率的に利用
- DPDK
 - 高速・安定なパケットフォワーディング
- Pluginでの機能拡張
 - Switching, Routing, ACL, NAT, SRv6, LB, etc
- 各種連携
 - ODL, OPNFV, OpenStack, k8s, etc
- CSIT: 機能・性能をコミュニティ内のCIで担保



- 13:30-13:55 FD.ioでCloud Nativeを加速 インテル 森直之氏&ユー・ピン氏
- インテルブースにてVPPの動態展示

ヤフーのL4 LBaaSの要求仕様と実装

要求仕様

- OSSで高速なデータプレーン
- L3DSR LB (using IPv4 DSCP field)
- Scaling-in/out capability
- CLOS上でECMP/BGPと連携

ヤフーのL4 LBaaSの要求仕様と実装

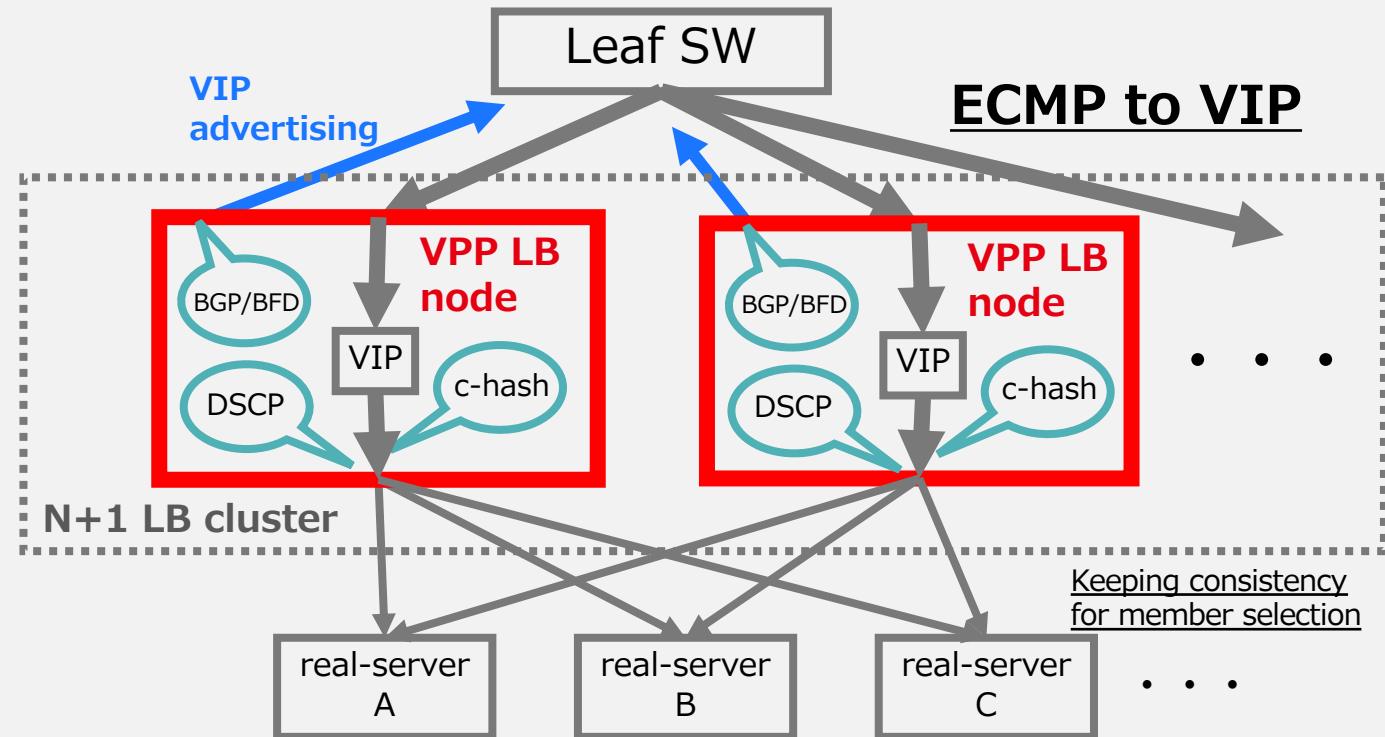
要求仕様 / 実装

- OSSで高速なデータプレーン
 - VPP
- L3DSR LB (using IPv4 DSCP field)
 - VPP LB plugins
- Scaling-in/out capability
 - Consistent-hash feature
- CLOS上でECMP/BGPと連携
 - Router pluginを利用

ヤフーのL4 LBaaSの要求仕様と実装

要求仕様 / 実装

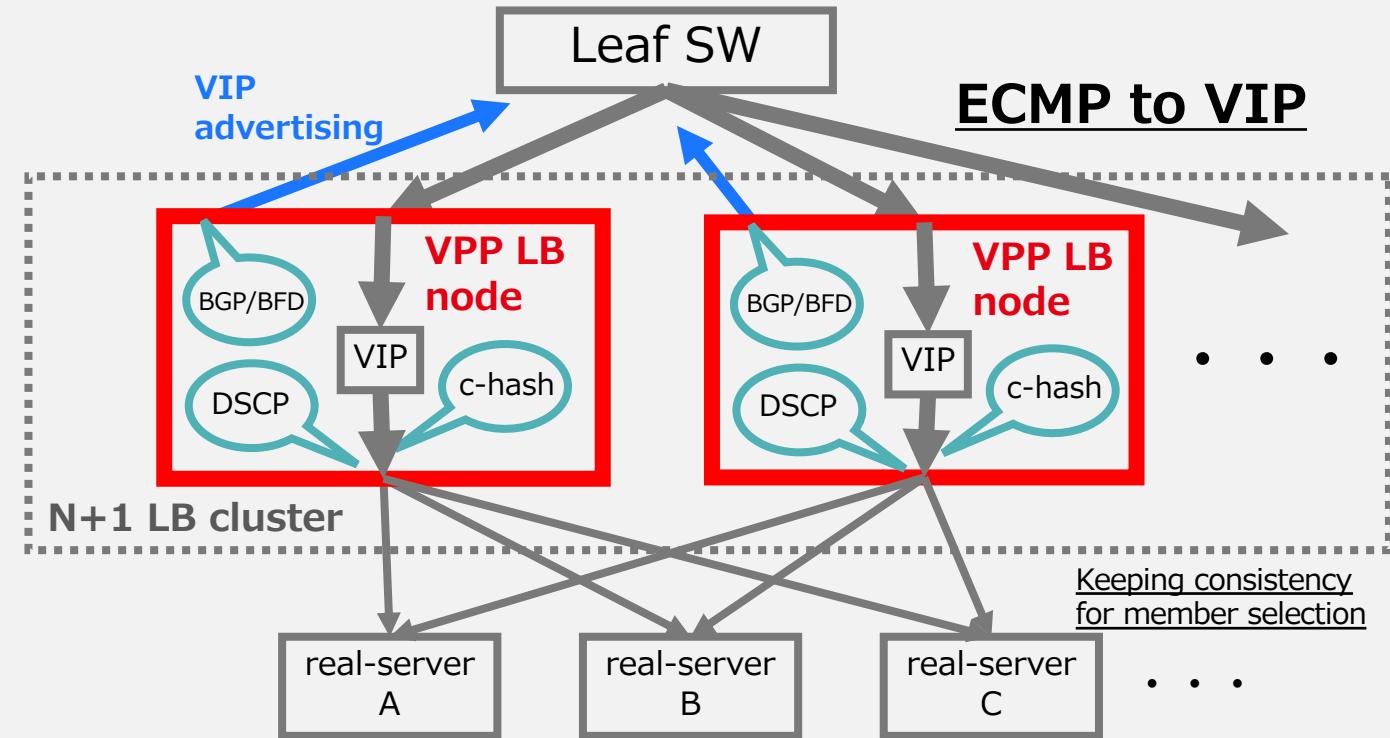
- OSSで高速なデータプレーン
 - VPP
- L3DSR LB (using IPv4 DSCP field)
 - VPP LB plugins
- Scaling-in/out capability
 - Consistent-hash feature
- CLOS上でECMP/BGPと連携
 - Router pluginを利用



ヤフーのL4 LBaaSの要求仕様と実装

要求仕様 / 実装

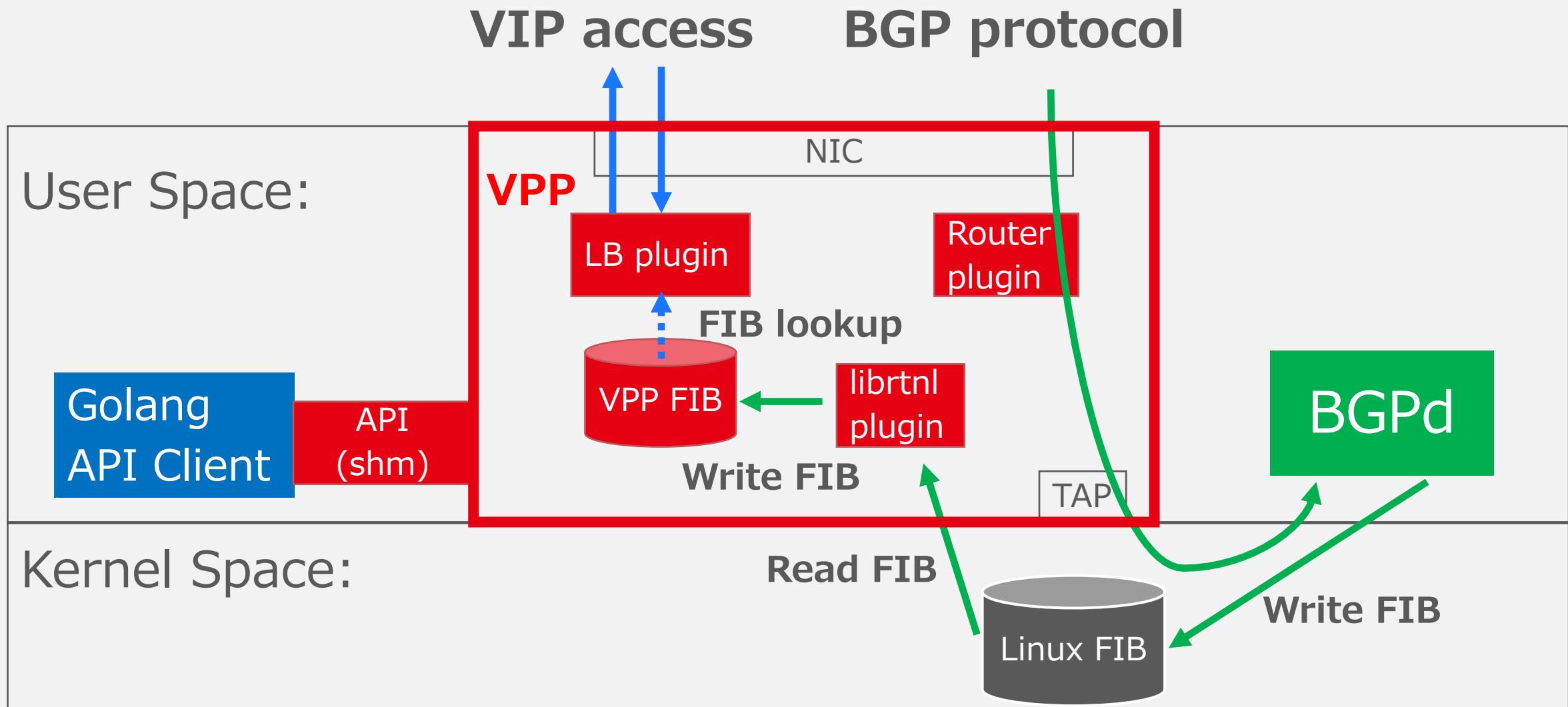
- OSSで高速なデータプレーン
 - VPP
- L3DSR LB (using IPv4 DSCP field)
 - VPP LB plugins
- Scaling-in/out capability
 - Consistent-hash feature
- CLOS上でECMP/BGPと連携
 - Router pluginを利用



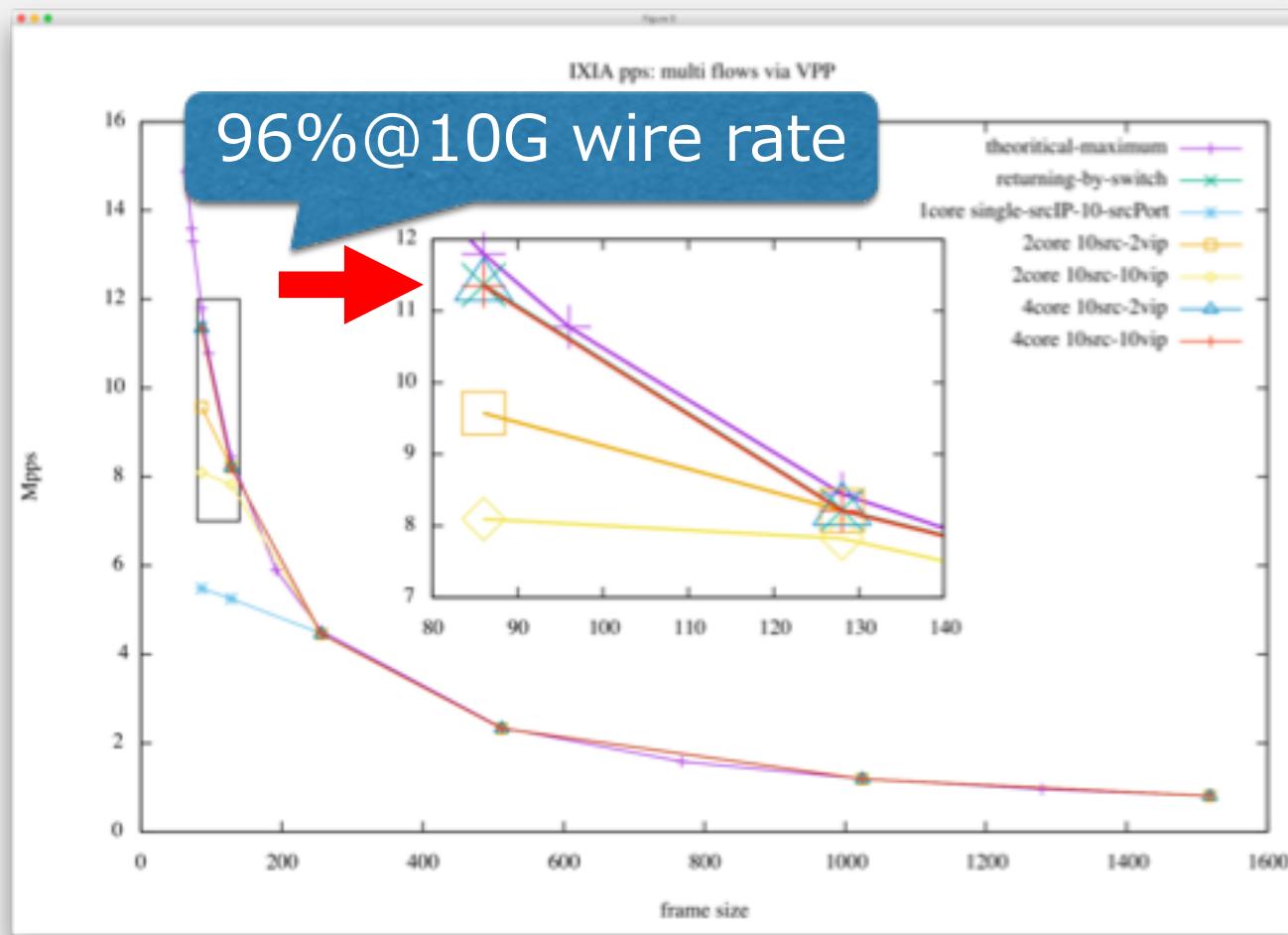
VPPで要求を実現:

- OSS based stateless L3DSR L4 Load Balancer
- Stable scaling-in/out N+1 capability
- CLOS ready, easy EoL (life cycle) management

VPPLB内部アーキテクチャ概要

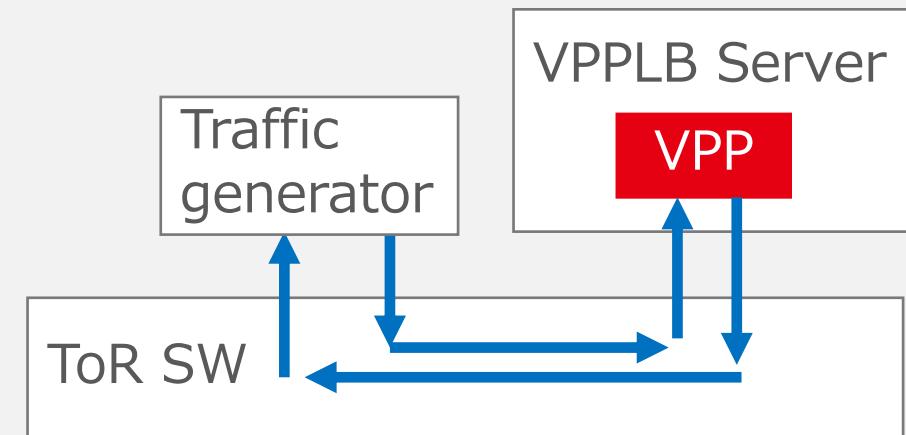


一台あたりのVPPLBスループット: 10Gbps



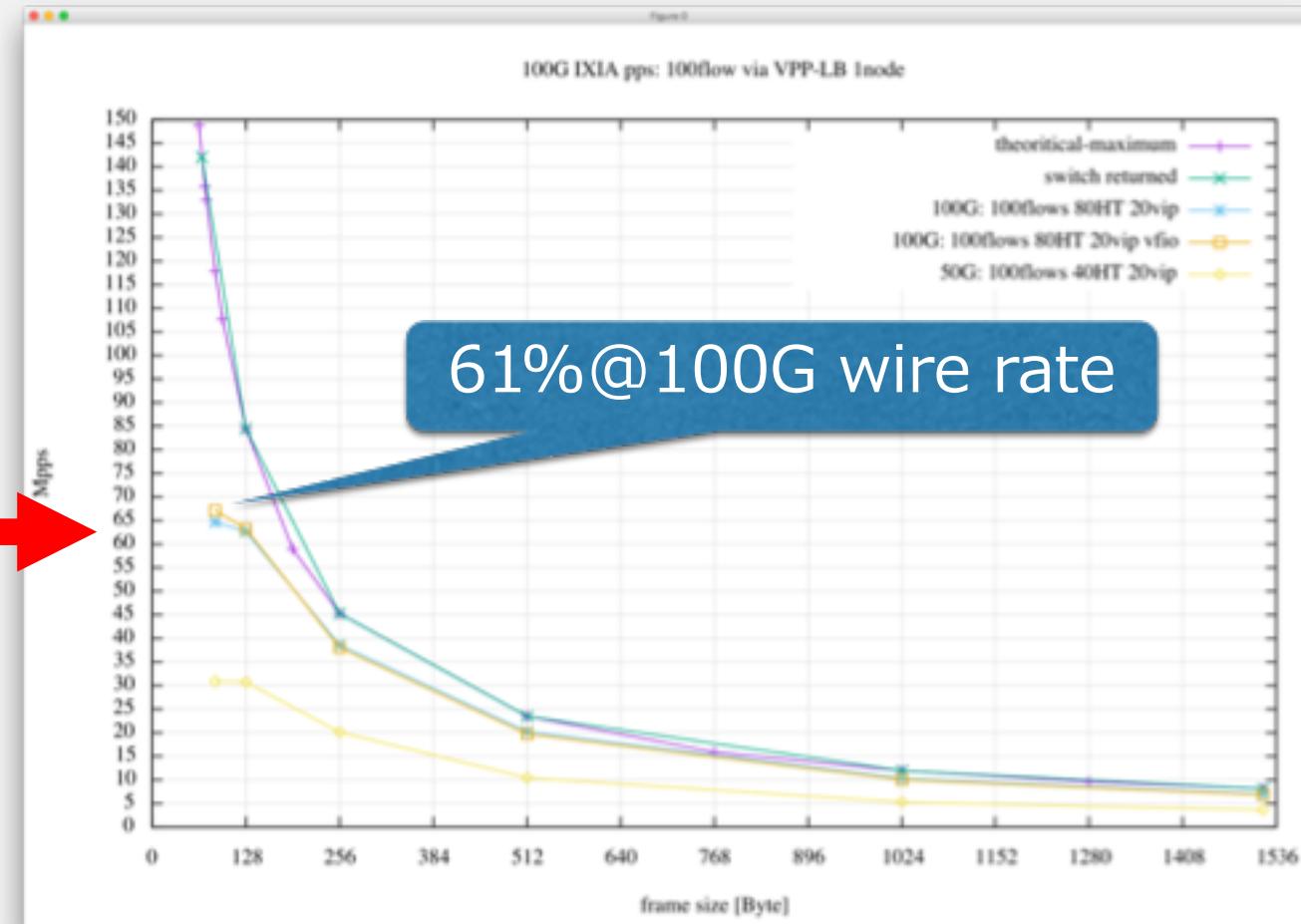
Traffic generator: Ixia
VPP Server Spec:

- CPU: Xeon E5-2650L v3 * 2
- Memory: 384GB
- NIC: Intel X540-AT2



10G環境で十分なパフォーマンスを実現！

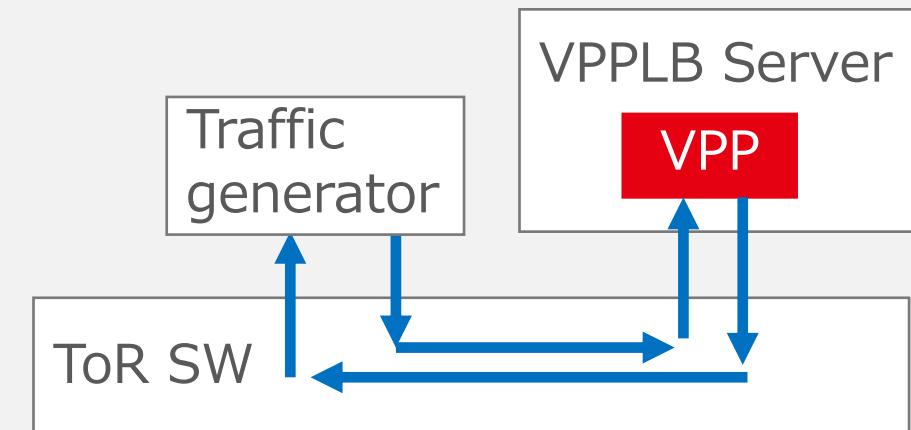
一台あたりのVPPLBスループット: 100Gbps



Traffic generator: Ixia

VPP Server Spec:

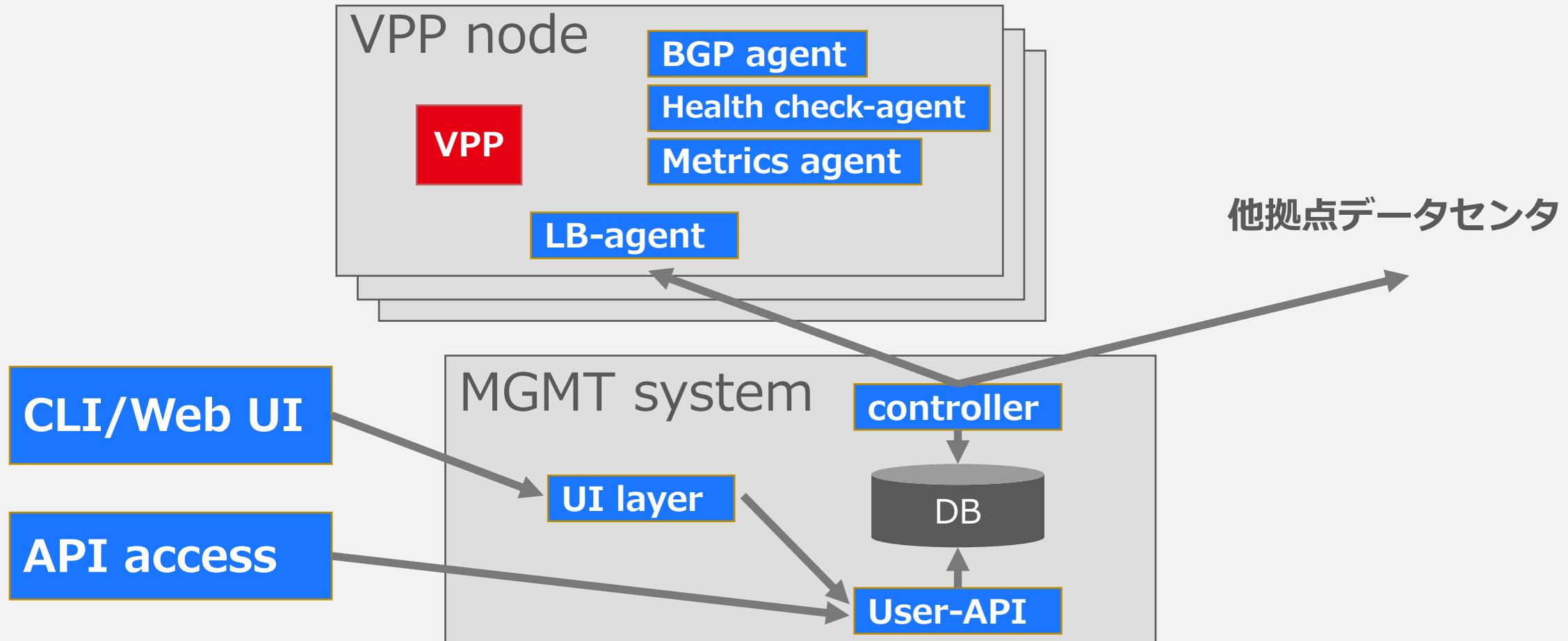
- CPU: Xeon Gold 6152 * 2
- Memory: 754GB
- NIC: Intel XXV710-DA2 * 2



次期フェーズに向けて100Gへの挑戦

LBaaS管理システム

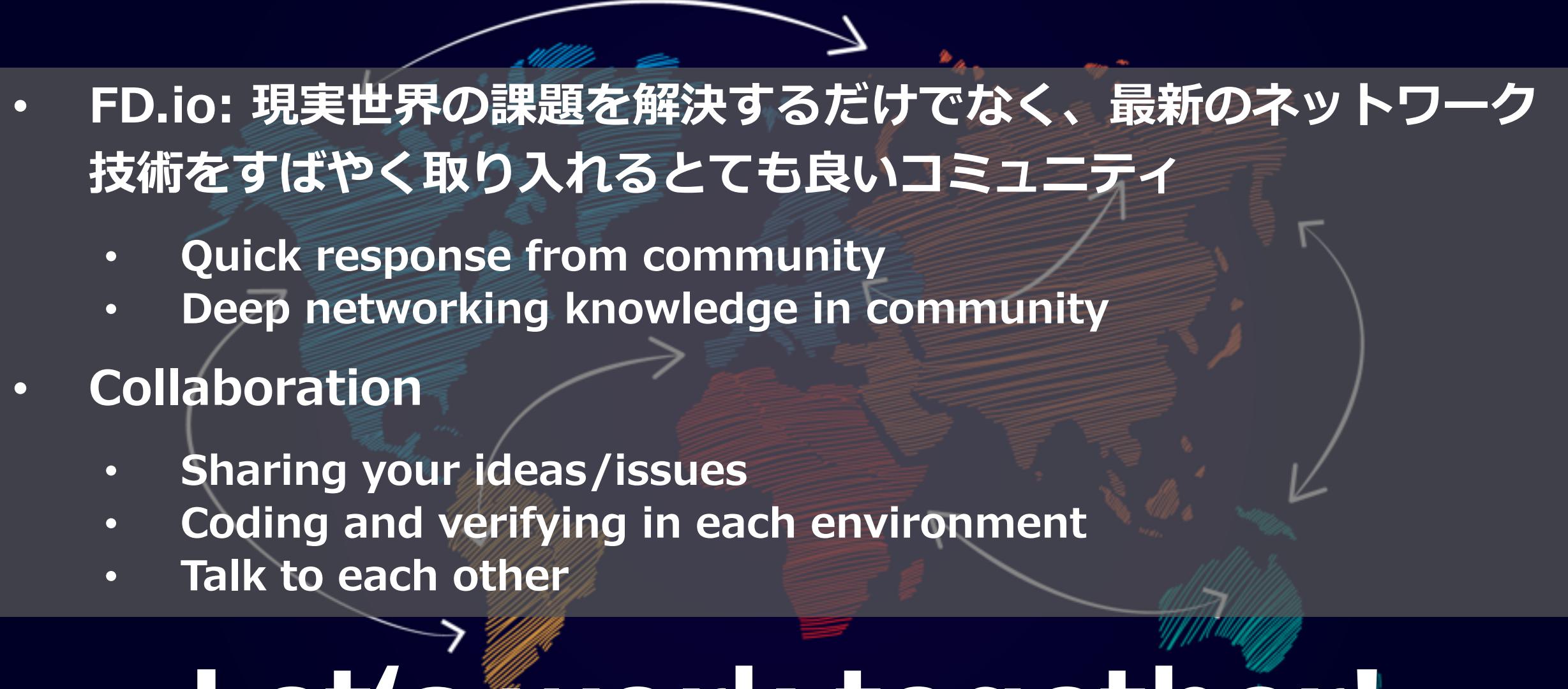
- 独自開発したVPP LBaaS管理システム



高速・安定したVIPデプロイを実現

Call to Action

- FD.io: 現実世界の課題を解決するだけでなく、最新のネットワーク技術をすばやく取り入れるとても良いコミュニティ
 - Quick response from community
 - Deep networking knowledge in community
- Collaboration
 - Sharing your ideas/issues
 - Coding and verifying in each environment
 - Talk to each other



Let's work together!

**Thank you
for your listening!**

