

自作ロードバランサ開発

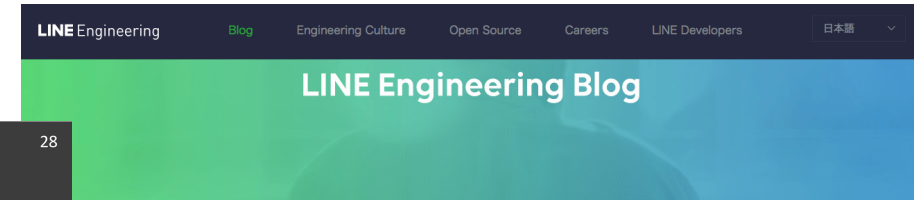
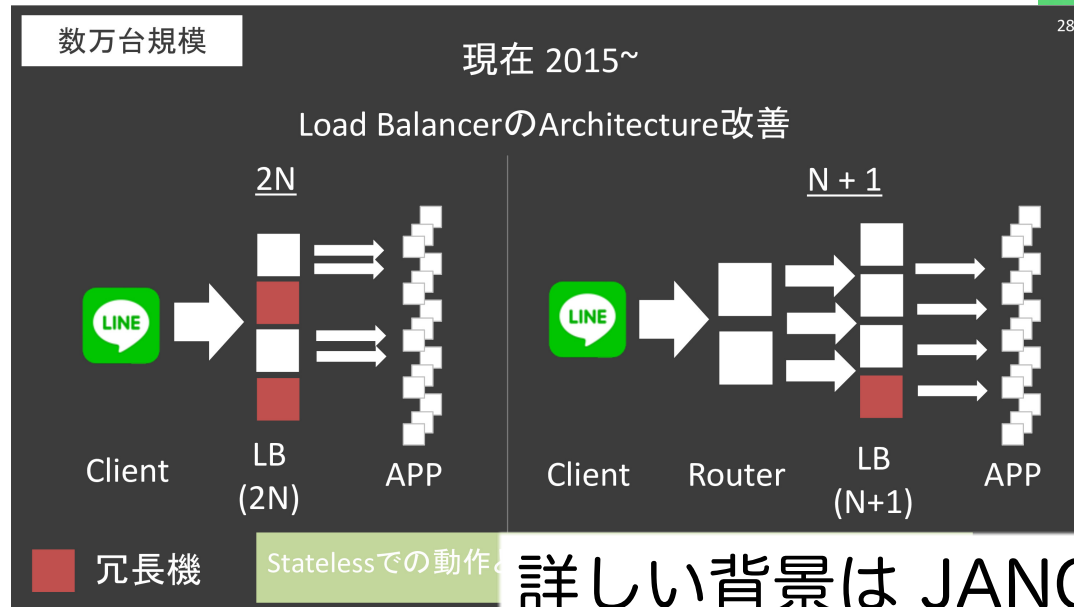
Yohei Kanemaru

LINE Corporation

yohei.kanemaru at linecorp.com / @yunazuno

こんなソフトウェアLBを開発しています

- LINEのサービスに必要な機能を備えている
- 普通のx86 Linuxサーバ上で動作する
- 高いスケーラビリティがある



JANOG39登壇レポート「LINEのインフラを運用して見えてきた課題」

saegusa | 2017.04.17

こんにちは。LINEでネットワークやデータセンターを担当している三枝です。2017年1月にJANOG39で登壇する機会を頂きましたので、今回はその内容をレポートしたいと思います。テーマはLINEのインフラについてです。LINEのサービス開始から現在までインフラの課題とその対応についてお話ししました。



詳しい背景は JANOG39 発表資料 & 登壇レポートにて

[1] LINEのインフラを運用して見えてきた課題 (JANOG39 発表資料) <https://www.janog.gr.jp/meeting/janog39/program/line.html>

[2] JANOG39登壇レポート「LINEのインフラを運用して見えてきた課題」 <https://engineering.linecorp.com/ja/blog/detail/135>

ロードバランサのアーキテクチャいろいろ

1. Centralized Load Balancers

Microsoft Ananta, Facebook shiv, Google Maglev, ...

2. De-centralized Load Balancers

GCP Internal LB, Cloudflare PoPs, ...

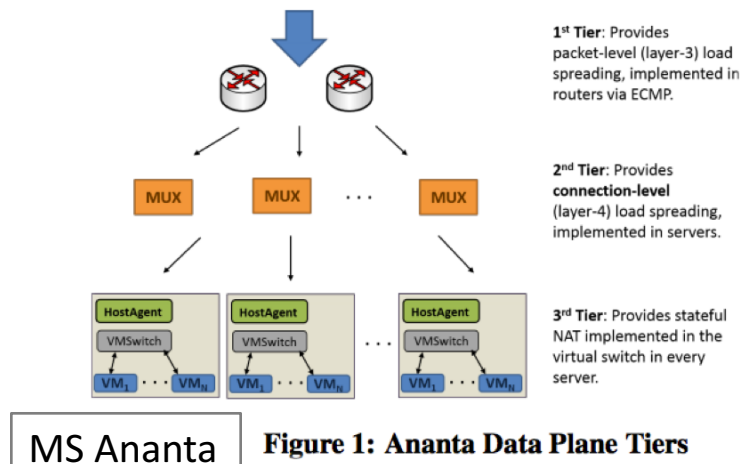


Figure 1: Ananta Data Plane Tiers

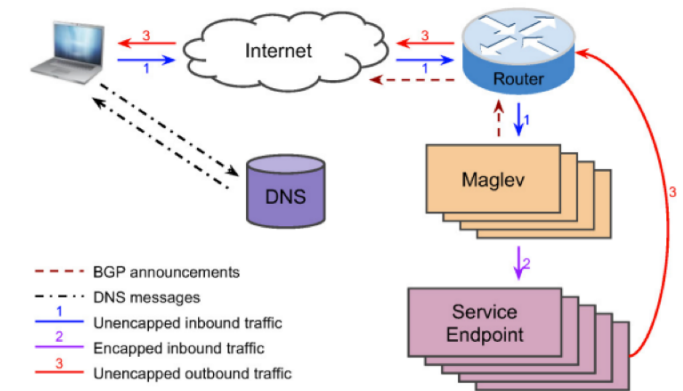
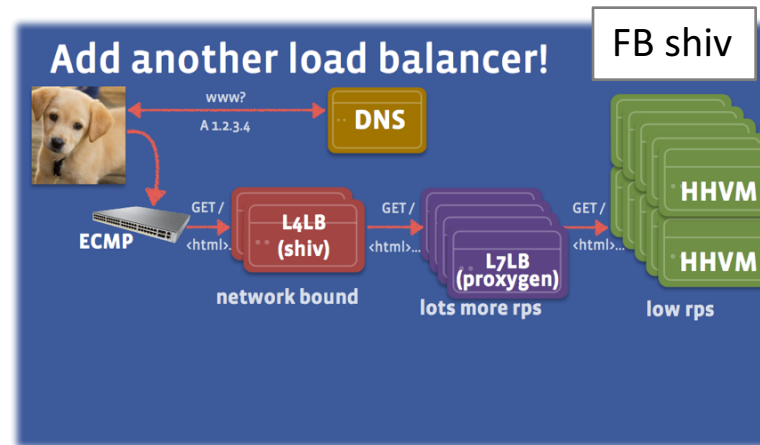


Figure 2: Maglev packet

Google Maglev

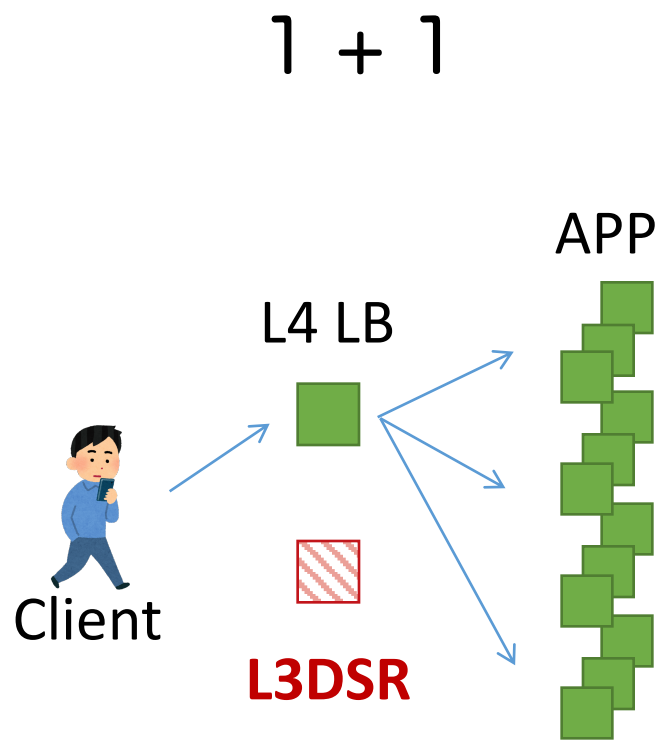
キーワード: N+1 & Multi-Tier

[1] Microsoft Ananta <https://doi.org/10.1145/2534169.2486026>

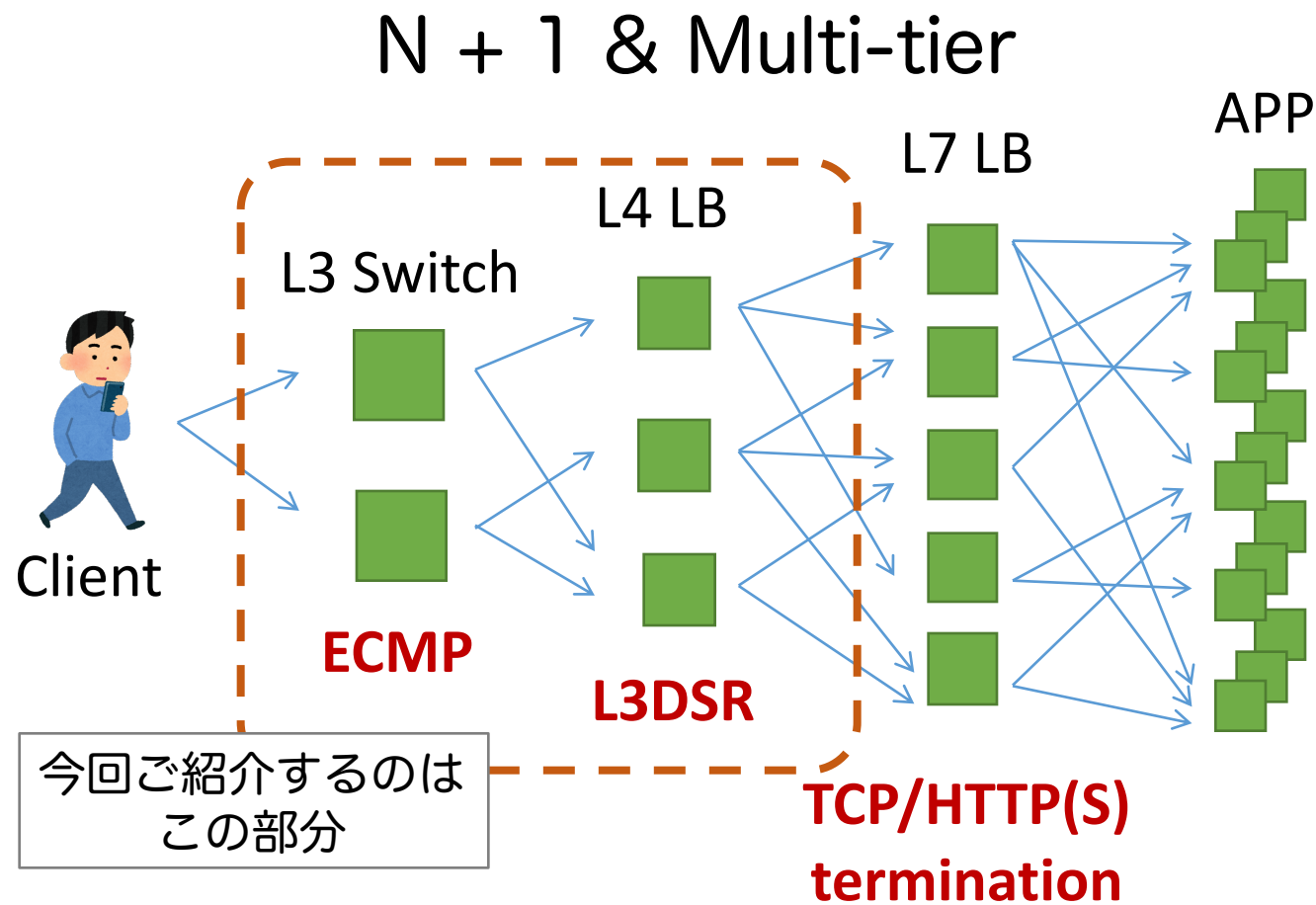
[2] Facebook Shiv: <https://www.usenix.org/conference/srecon15/program/presentation/shuff>

[3] Google Maglev <https://research.google.com/pubs/pub44824.html>

1+1構成から N+1 & Multi-tier構成へ

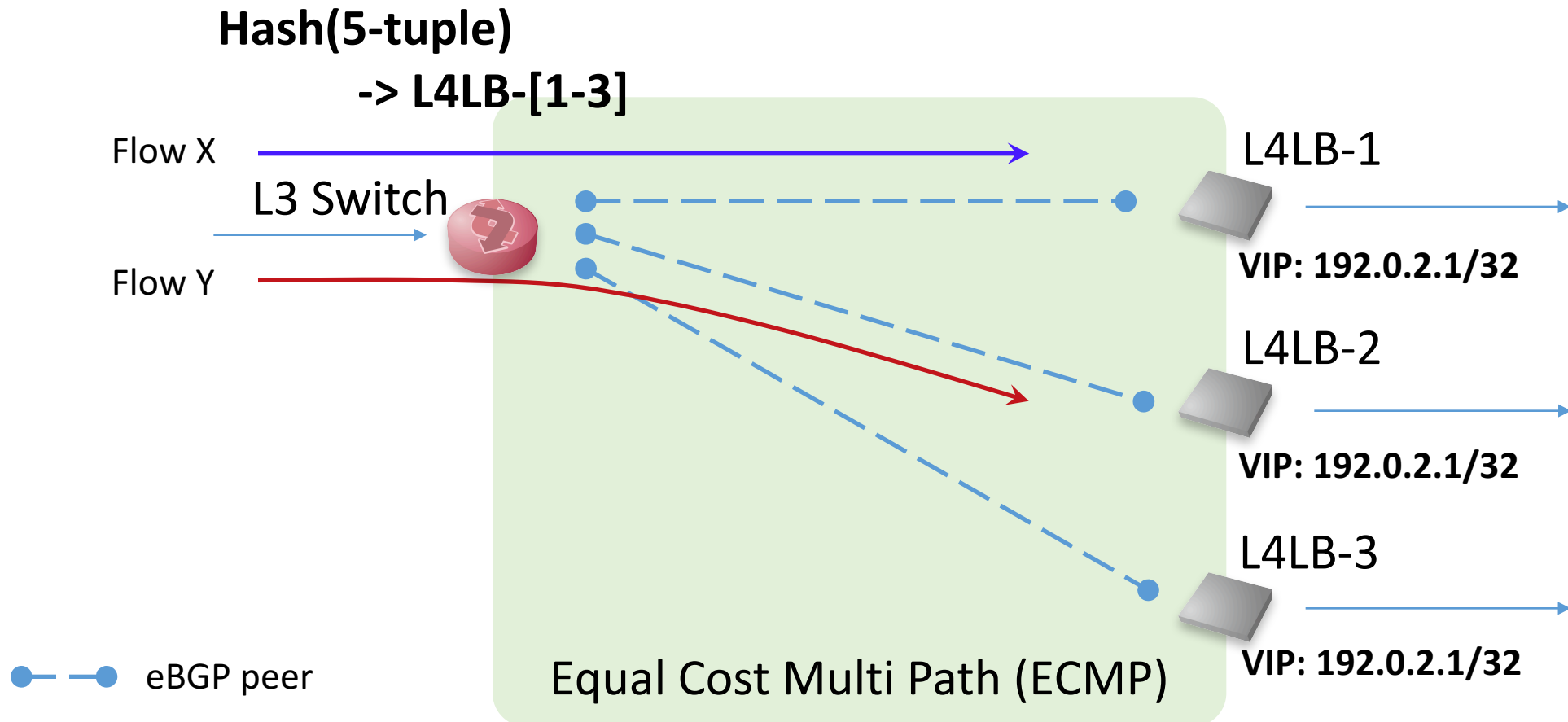


■ : In service
▨ : Not in service



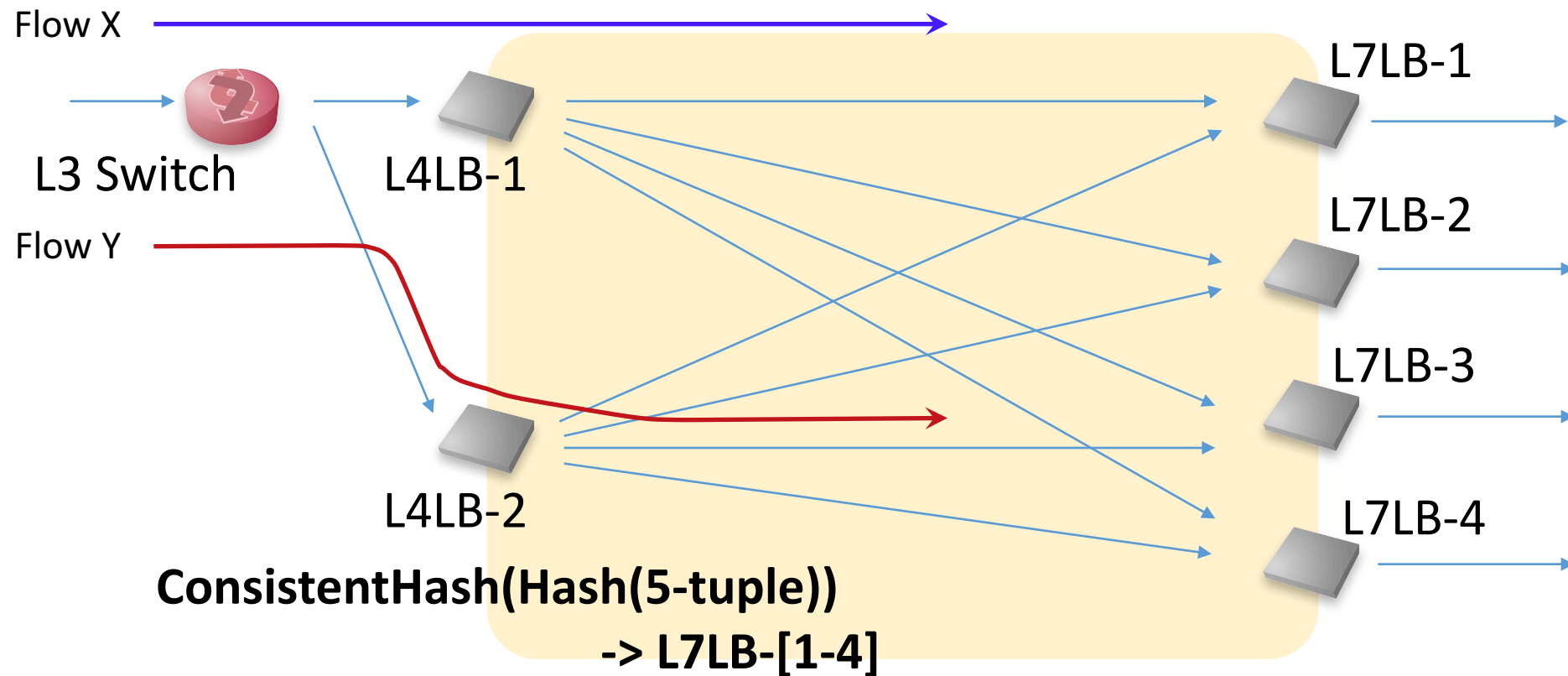
1段目 :: BGP ECMP

Per-flow ECMP により、L4LBへおおよそ均等にトラフィックを分散



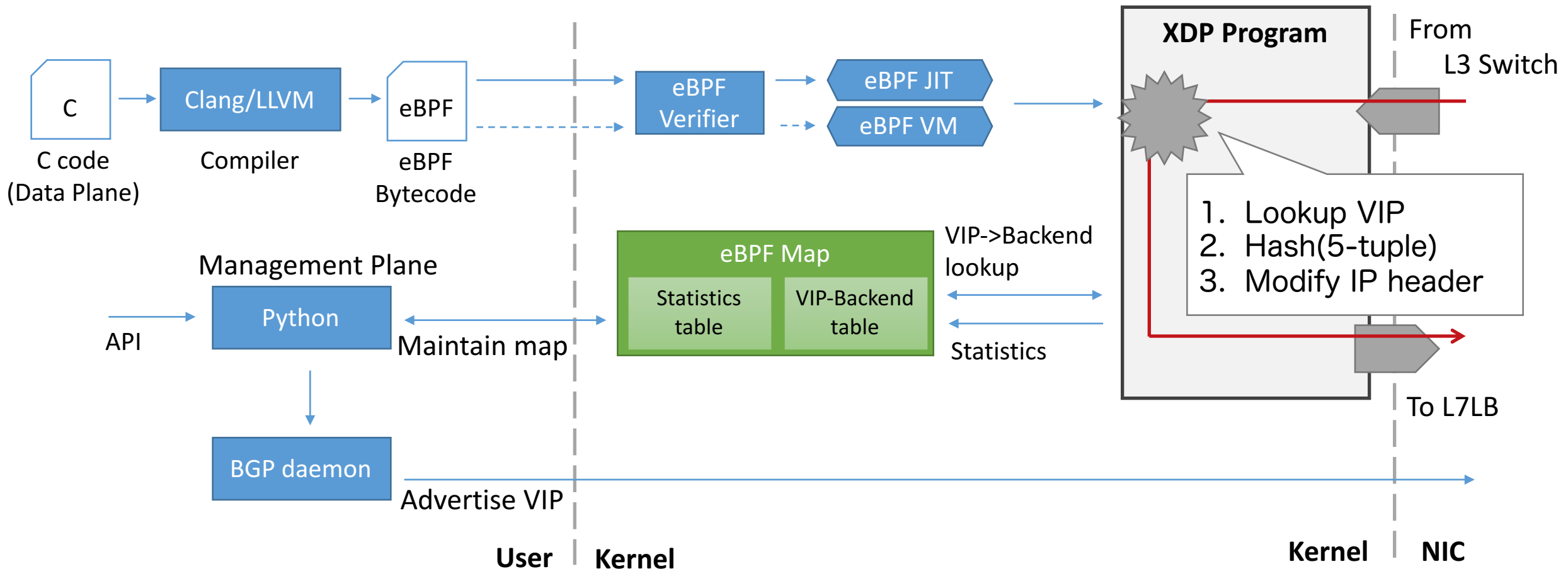
2段目 :: Stateless L4LB

Consistent Hashで分散先L7LBを決定し、L3DSRでパケット転送

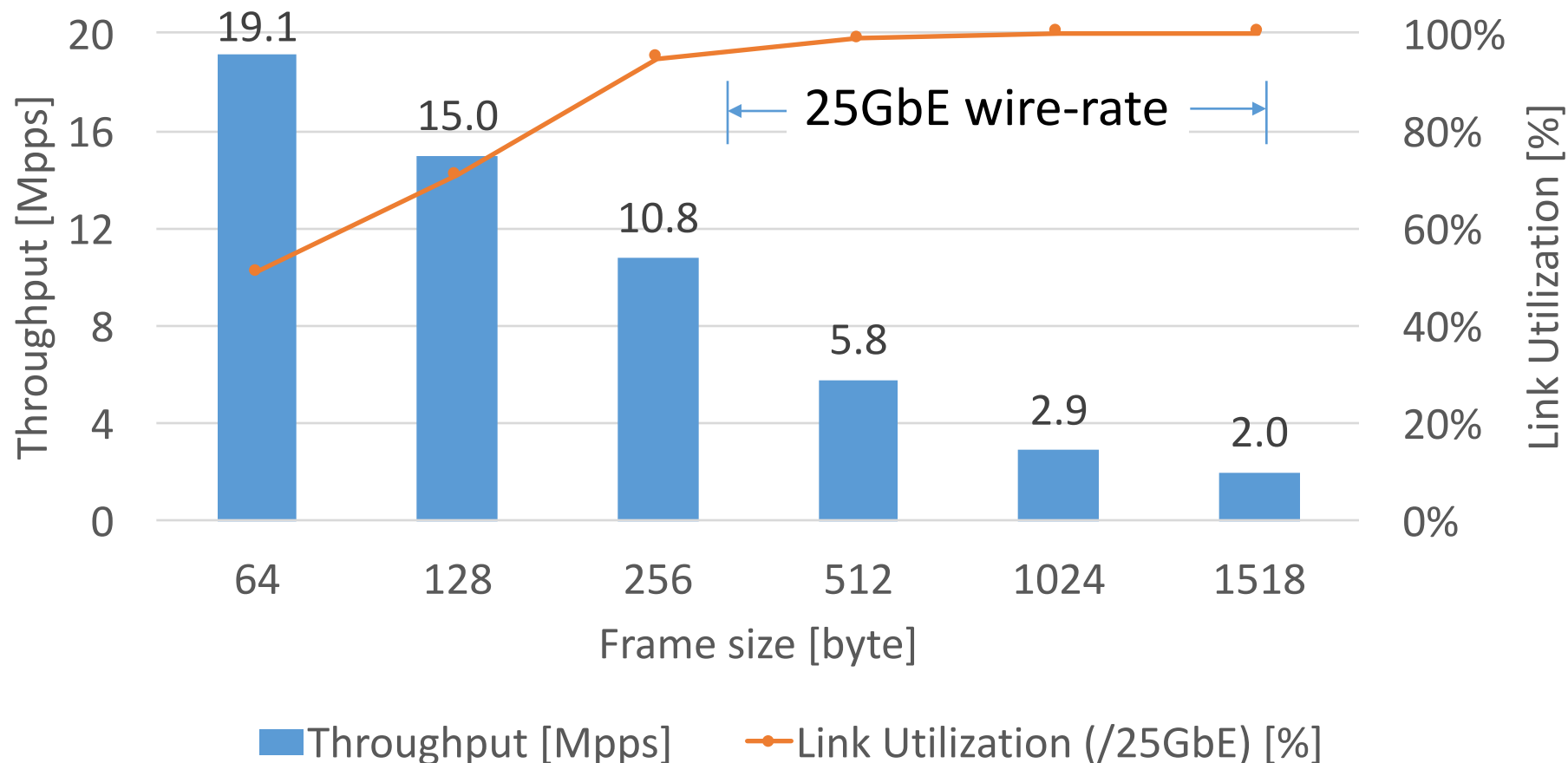


L4LB :: eXpress Data Path (XDP) を活用

Linux kernel内で動作する、eBPFを用いた高速パケット処理基盤



L4LB :: ノード単体パフォーマンス



Environment

Hardware: 2 x Intel Xeon E5-2630 v4 (10 cores) / 32G DDR4-2133 / Mellanox ConnectX-4 Lx EN 25GbE

Software: Linux Kernel 4.11.0-rc5+[dcb421f4](#) / BPF JIT enabled

実際のところどんな状況？

当初の狙いはおおむね達成できそう

手元のコードと同一のものが実ネットワークでそのまま動作する
テストがしやすい・動作が見えやすい

N+1 構成が問題になるケースもある

Multipath TCP, IP Fragment, 一部の特殊案件, ...

今後の展望

本格的にプロダクション環境への投入を開始

機能追加や構成のアップデート

- DoS / DDoS 防御

- ステートフルセッションキャッシュ (LRU cache) の併用

- BGP unnumbered (RFC 5549)

- DC networkのL3化

ハードウェア寄りのアプローチを調査・検討

- XDP (eBPF) hardware offload, Programmable ASIC, P4, ...?