CHINA OpenInfra Days

CHINA OpenInfra Days | IT大咖说 知识共享平台

# 容器在无线网络中的性能分析实践

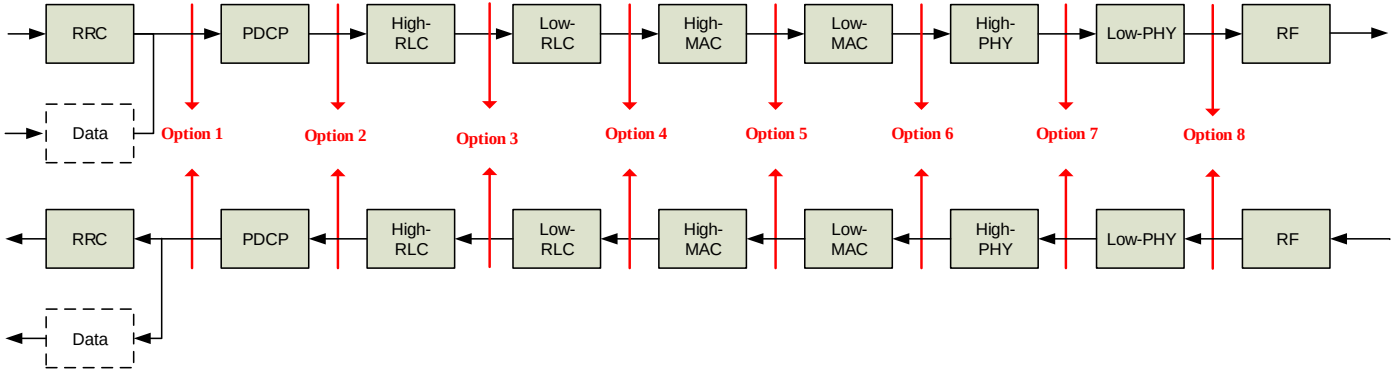
李贤明 博士 华为预研工程师
李明东 华为预研工程师
黄骋 华为预研工程师

# 提纲

- 背景
- 评测系统
- 评测实践
- 标准进展
- 未来工作

# 背景

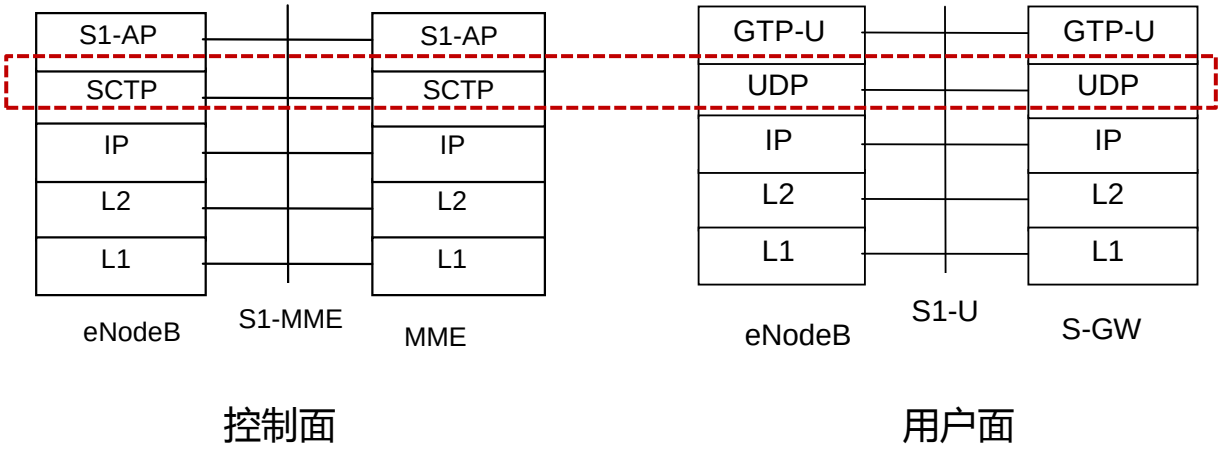
- 3GPP RAN演进




- 5G NR的切分存在多种备选方式，2017年3GPP对RAN架构高层切分达成一致结论：Option 2
- RLC、MAC、PHY和RF归属DU，RRC、PDCP则归属CU，PDCP作为可独立部署或伸缩的单元。

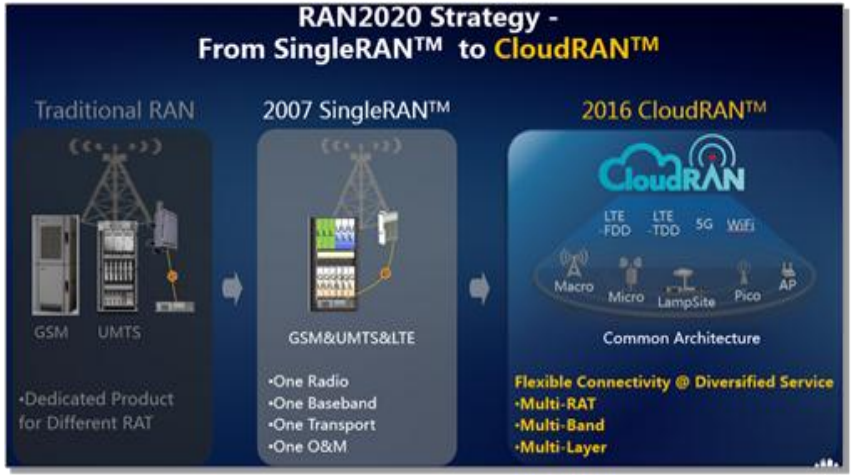
# 背景

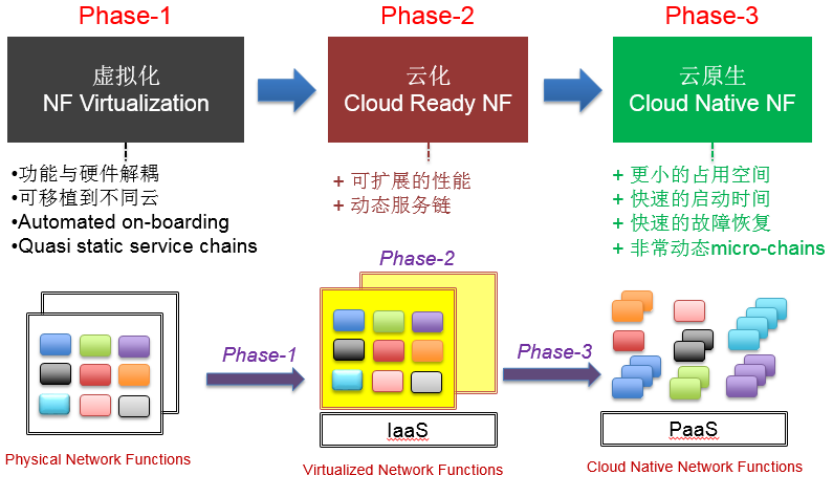
- 4G RAN接口协议栈(23.401)



# 背景

- 华为CloudRAN



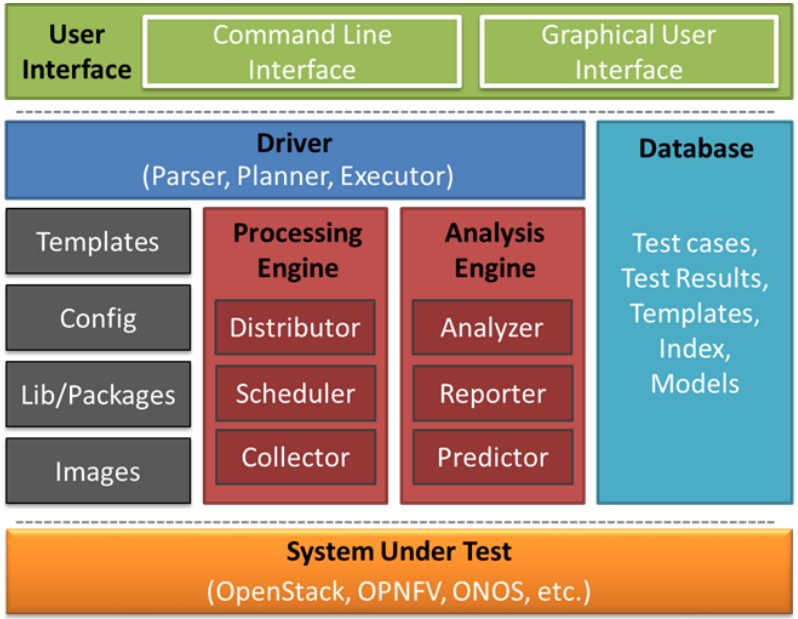

- 基站形态：分布式 → 集中式 → 分布+集中
- 云化阶段：Virtualization → Cloudification → Cloud Native ?

# 背景

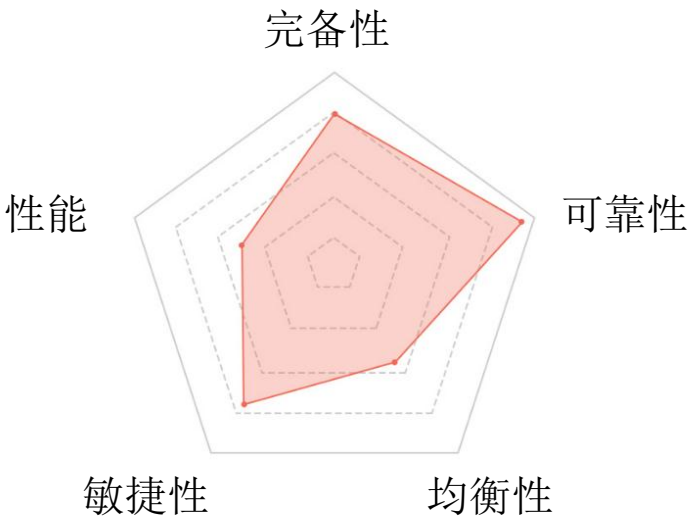
- 评测需求
  - 面对Cloud Native的趋势，传统VM or Container？
  - K8S的编排性能能否满足CloudRAN业务的要求？
  - CloudRAN云化业务的衡量指标有哪些？影响其性能的因素又该如何评价？

# 评测系统

- WIPM(Wireless Infrastructure Performance Modeling) 1.0 **2017.6**




自动化测试系统

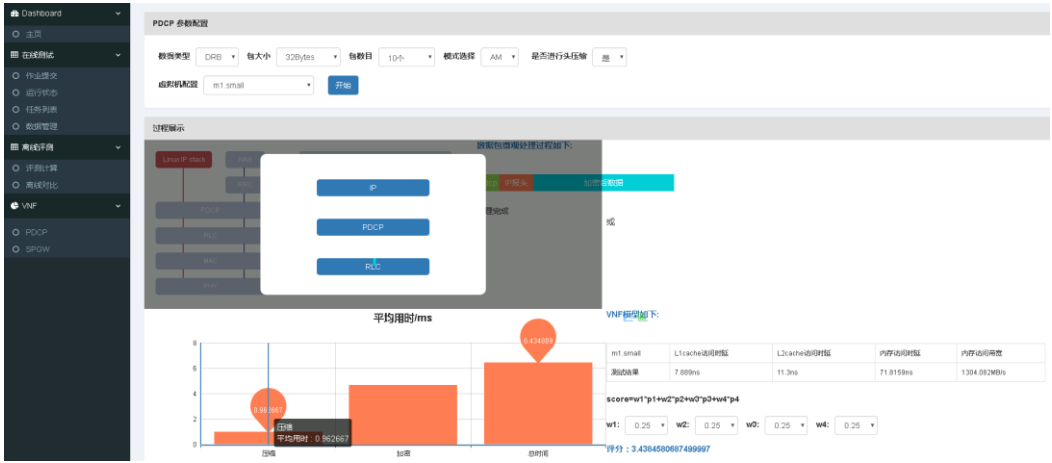



五维评估模型

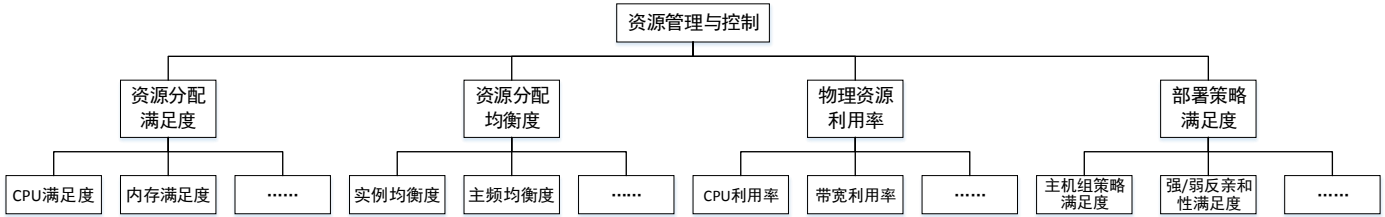
# 评测系统

- WIPM 2.0 **2018.5**


移动端展示




VNF评测功能增强，目前支持PDCP模块时延测试，和MME–SPGW的发包测试




测试指标与测试用例扩充

# 评测实践

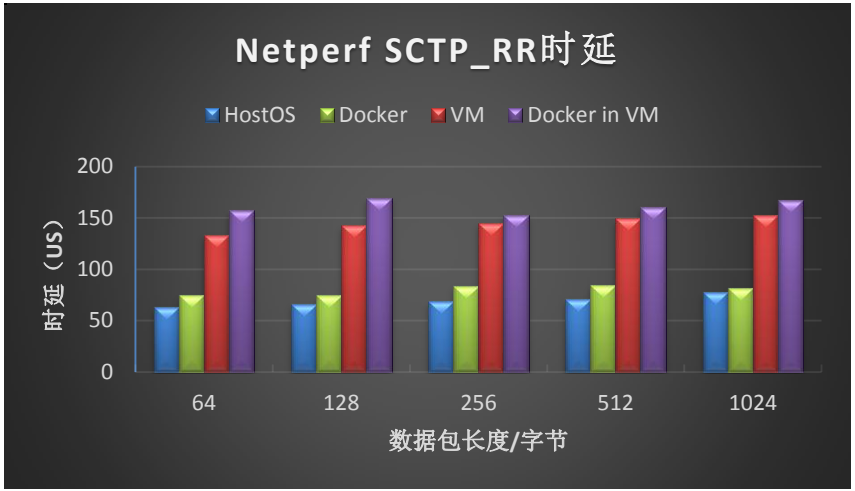
- 测试环境

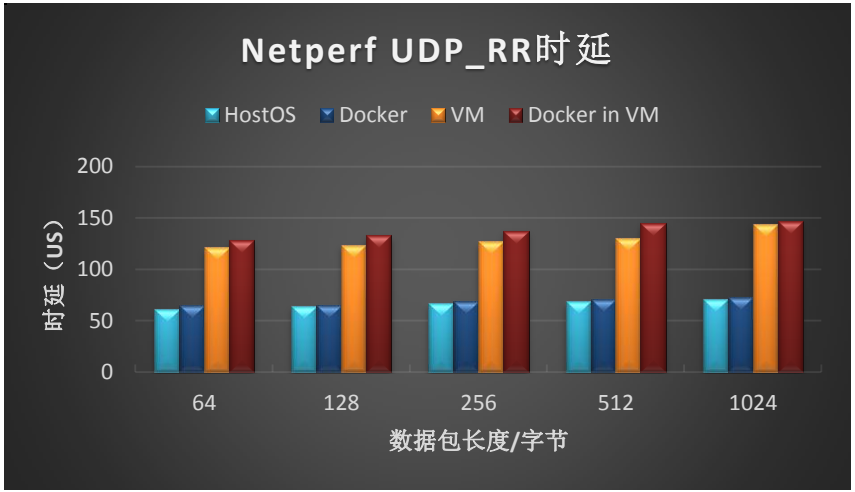
| 硬件环境 | | | |
|---|---|---|---|
| 服务器 | | | |
| CPU型号 | Intel Xeon E7-4850 v2 @ 2.30GHz | Intel Xeon E5-2650 @ 2.00GHz | Intel Xeon X5670 @ 2.93GHz |
| CPU个数 | 4 | 2 | 2 |
| 内存 | 512G | 128G | 80G |
| 硬盘 | 600G*8 | 2.2T | 300G |
| 网卡 | Intel 82599 | Intel 82580 | Intel 82599 |
| 网卡个数 | 2 | 1 | 1 |
| 交换机 | | | |
| 网口 | 24个10/100/1000Base-T<br>4个100/1000 Base-X 千兆Combo口 | | |

| 软件环境 | |
|---|---|
| Ubuntu | 16.04 |
| OPNFV | Danube 3.0（基于OpenStack Newton） |
| Kubernetes | 1.10 |
| QEMU-KVM | 2.5.0 |
| OVS | 2.6.1 |
| pktgen-dpdk | 3.4.2 |
| MoonGen | commit 31af6e66eae4fd77a399cc0d7be555f3b8132d7a |
| Docker | 1.13.1，build 092cba3 |
| DPDK | 17.05 |

# 评测实践

- 控制面和用户面时延

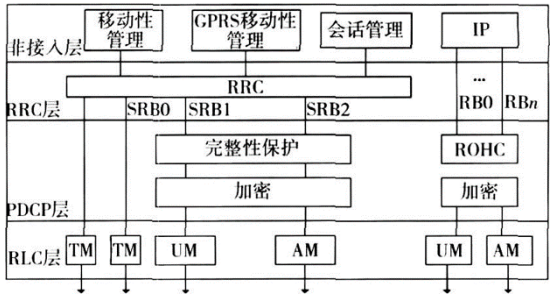


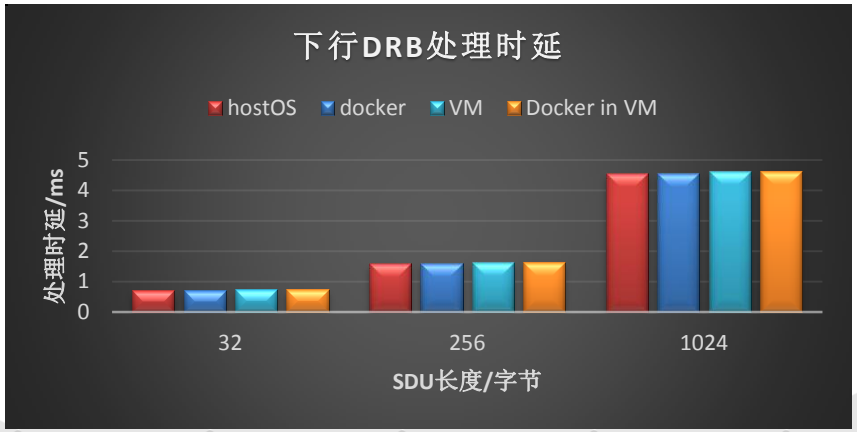

- 容器相比于宿主环境性能损失较小，控制面SCTP和用户面UDP时延分别增加18%和12%
- 裸机容器相比VM时延降低40%以上，更适合低时延业务的部署

# 评测实践

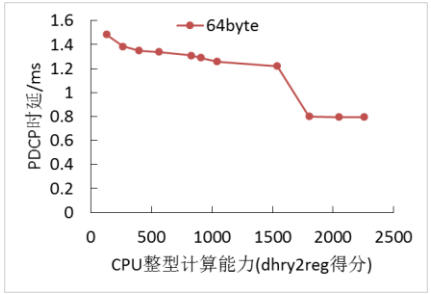
- 基于OAI的PDCP性能分析

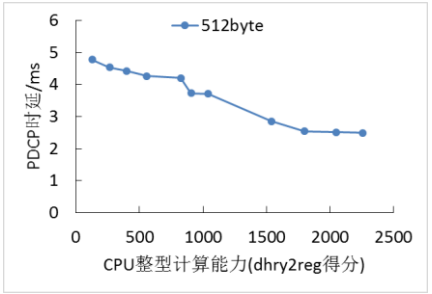



```
├── openair2 :3GPP LTE Rel-10 implementation.
    ├── LAYER2/RLC/
    ├── LAYER2/PDCP
    ├── RRC/LITE
    ├── PHY_INTERFACE
    ├── X2AP
    ├── ENB_APP
```




- docker create -it --cpu-period=100000 --cpu-quota=50000 --name pdcp_0.5 pdcp-test
- docker exec pdcp_0.5 /root/test_pdcp 32 DRB am




- 对于计算密集型的PDCP，VM、裸机容器和容器内VM计算性能差别不大
- 下行DRB数据包处理时延随CPU整型计算能力的增强单调下降，但并不存在严格的线性关系

# 评测实践

- PDCP性能建模 — 特征选择
  - 衡量数据集中每个特征对性能KPI的影响程度
    - ➢计算第i个特征$X_i$与性能$y$的相关系数

$$r_i = \frac{(X_i - \bar{X}_i)'(y - \bar{y})}{Std(X_i)Std((y)}$$

    - ➢计算F值

$$F = \frac{r_i^2}{1 - r_i^2}(n - 2)$$

  - 选择对性能KPI影响最大的top k关键因素
    - ➢k给定：选择F值最大的k个特征
    - ➢k待定：将特征按F值排序，顺序遍历，F值相差2个数量级时截断

| 单核双精度计算能力 | 单核字符串处理能力 | 多核双精度计算能力 | 多核字符串处理能力 | 内存访问带宽 | 内存访问时延 | 磁盘随机读速度 | 磁盘随机写速度 | DRB长度 | PDCP处理时延 |
|---|---|---|---|---|---|---|---|---|---|
| 0.042 | 0.036 | 0.003 | 0 | 0.240 | 0.060 | 0.027 | 0.027 | 0 | 0.653 |
| 0.054 | 0.063 | 0.011 | 0.016 | 1 | 0 | 0.001 | 0.001 | 0 | 0.535 |
| …… | | | | | | | | | |

计算F

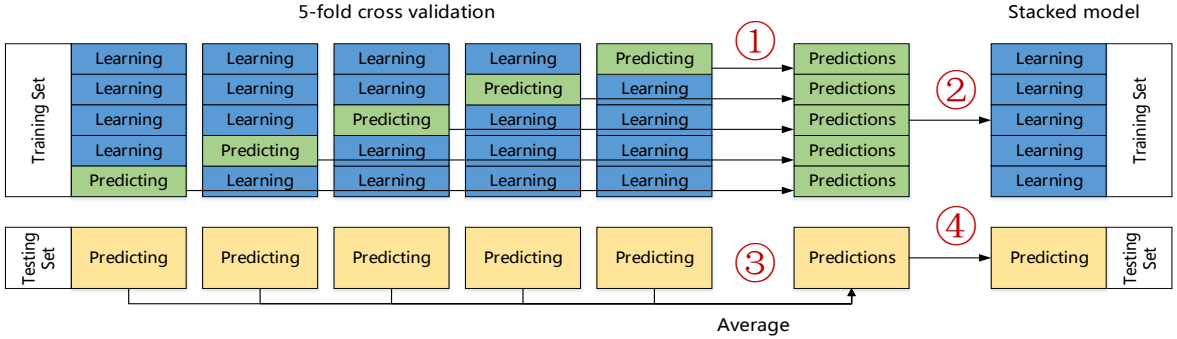
| | DRB长度 | 单核字符串处理能力 | 多核字符串处理能力 | 单核双精度计算能力 | 多核双精度计算能力 | 磁盘随机写速度 | 磁盘随机读速度 | 内存访问时延 | 内存访问带宽 |
|---|---|---|---|---|---|---|---|---|---|
| F值 | 3289 | 3.96 | 3.64 | 3.53 | 2.01 | 1.70 | 1.66 | 0.52 | 0.006 |

k = 5

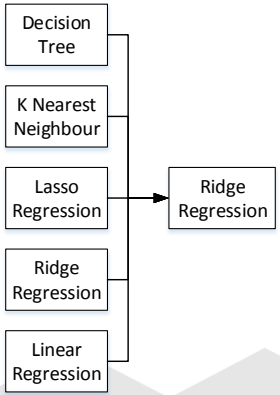
| 单核双精度计算能力 | 单核字符串处理能力 | 多核双精度计算能力 | 多核字符串处理能力 | DRB长度 | PDCP处理时延 |
|---|---|---|---|---|---|
| 0.042 | 0.036 | 0.003 | 0 | 0 | 0.653 |
| 0.054 | 0.063 | 0.011 | 0.016 | 0 | 0.535 |
| …… | | | | | |

# 评测实践

- PDCP性能建模 — 性能预测（Stacking）




①：通过5折交叉验证对每个基预测器进行训练
②：将不同基预测器的预测数据汇总成新数据集，在此基础之上训练次级预测器
③：获取每个基预测器在测试集上的预测结果
④：通过次级预测器获得预测性能




PDCP时延预测平均相对误差(threshold:10%)

| 训练集 | 测试集 | Linear Regression | Ridge Regression | Lasso Regression | K Nearest Neighbour | Decision Tree | Stacking |
|---|---|---|---|---|---|---|---|
| 2288,2285,2485 | 5885 | 0.544 | 0.329 | 0.279 | 0.383 | 0.256 | 0.254 |
| 5885,2285,2485 | 2288 | 0.411 | 0.356 | 0.383 | 0.236 | 0.184 | 0.183 |
| 5885,2288,2485 | 2285 | 3.975 | 2.399 | 0.603 | 0.350 | 0.311 | 0.311 |
| 5885,2288,2285 | 2485 | 0.712 | 0.462 | 0.462 | 0.251 | 0.157 | 0.166 |
| 混合数据集的75% | 混合数据集的25% | 0.568 | 0.527 | 0.369 | 0.398 | 0.187 | 0.186 |

# 评测实践

- K8S编排

| | Pod | CPU(核) | 内存(Gi) | 数量 |
|---|---|---|---|---|
| 平台网元 | A | 8 | 16 | 3 |
| | B | 6 | 16 | 2 |
| | C | 4 | 8 | 3 |
| | D | 4 | 8 | 2 |
| | E | 4 | 24 | 3 |
| | F | 4 | 16 | 2 |
| 业务网元 | G | 8 | 24 | 1 |
| | H | 8 | 24 | 1 |
| | I | 12 | 36 | 1 |

| | 节点1 | 节点2 | 节点3 | 节点4 |
|---|---|---|---|---|
| CPU(核) | 32 | 32 | 96 | 96 |

部署策略

1. 实例化1个平台网元
2. 实例化1个业务网元
3. 业务网元Scaling

| 业务网元副本数 | 已分配CPU核(容量占比) | | | |
|---|---|---|---|---|
| | 节点1 | 节点2 | 节点3 | 节点4 |
| 1 | 12(38%) | 12(38%) | 38(40%) | 42(44%) |
| 2 | 20(63%) | 12(38%) | 50(52%) | 50(52%) |
| 3 | 28(88%) | 24(75%) | 50(52%) | 58(61%) |
| 4 | 28(88%) | 24(75%) | 58(61%) | 78(81%) |
| 5 | 28(88%) | 24(75%) | 78(81%) | 86(90%) |
| 6(×) | 28(88%) | 24(75%) | 90(94%) | 94(98%) |

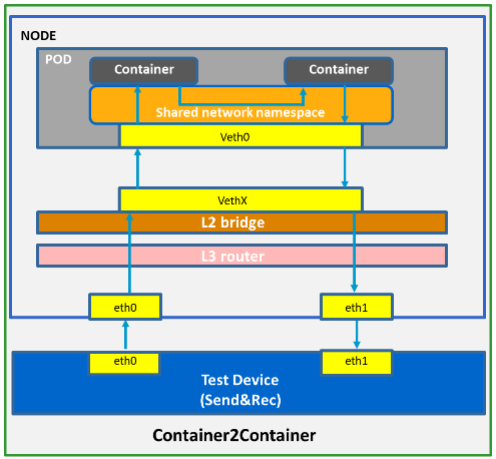
- K8S基于requested / capacity的打分策略能够保证资源分配的均衡性，但会造成资源碎片，无法最大化资源利用率
- 在对pod进行扩容操作时，K8S还不支持通过pod迁移整合资源碎片从而满足资源需求

# 评测实践

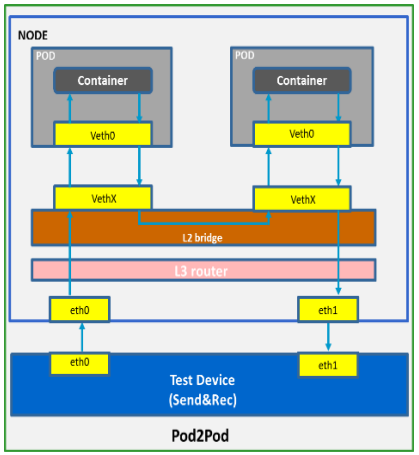
- 实践体会
  - 相比传统VM，容器在控制面和用户面时延方面性能优势明显，适合低时延NF的部署
  - 对于计算密集型的NF，容器与传统VM性能差别并不大，能够通过性能建模准确地预测其性能
  - 目前K8S的资源编排方案还不能完全满足云化无线业务场景的需求，需要在提高资源利用率和资源碎片的灵活整合方面做相应的增强

# 标准进展

- NFV TST 009(Networking Benchmarks and Measurement Methods for NFVI)贡献两个test setup




- 测试单个Pod内部容器之间的网络性能
- Use case：PDCP内部的加解密模块和转发模块的数据传输



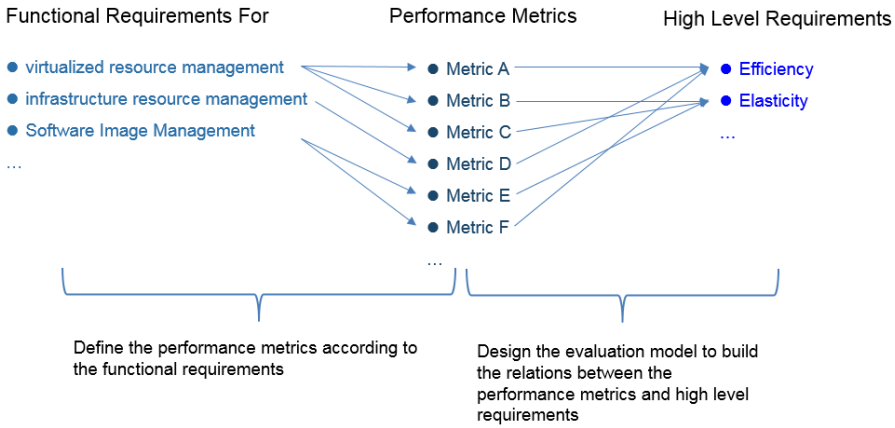

- 测试同Node上Pod和Pod之间的网络性能
- Use case：RRC和PDCP之间的信令传输

Container2Container

Pod2Pod

# 标准进展

- NFV TST NWI: Report on VIM&NFVI Control and Management Performance Evaluation（立项已获TST组内通过）




| | Performance Metric | Functional Requirements |
|---|---|---|
| Compute Resource Management | Time to create/delete/update/query a VM instance | IFA010 Vim.Vrm.001 |
| | Time to start/stop/pause/reboot a VM instance | IFA010 Vim.Vrm.001 |
| | Time to scale up/down a VM instance | IFA010 Vim.Vrm.001 |
| | Time to create an affinity/anti-affinity group | IFA010 Vim.Vrm.006 |
| Network Resource Management | Time to create an external/internal virtual network | IFA010 Vim.Nfpm.001 |
| | Time to create/delete a vSwitch/vRouter | IFA010 Vim.Nfpm.001 |
| | Time to create/delete a virtual subnet | IFA010 Vim.Nfpm.001 |
| Software Image Management | Time to create, delete update or query images | IFA010 Vim.Sim.001 |
| …… | …… | …… |

- 主要关注VIM资源管理能力和效率的评测
- IFA功能需求 — 性能测试指标 — 整体评价方法

# 未来工作

- RANCU性能建模
- NFV TST WI继续完善
- 面向Edge Cloud的性能评测

CHINA
OpenInfra Days

IT大咖说
知识共享平台

# Thank You