vrAIn

A Deep Learning Approach to Virtualized Radio Access Networks (vRAN)

Andres Garcia-Saavedra
NEC Laboratories Europe

Marco Gramaglia
Universidad Carlos III de Madrid

Ayala-Romero, J.A., Garcia-Saavedra, A., Gramaglia, M., Costa-Perez, X., Banchs, A. and Alcaraz, J.J., *vrAIn: A Deep Learning Approach Tailoring Computing and Radio Resources in Virtualized RANs.* ACM MobiCom 2019.

New requirements from the network management

- The introduction of novel networking paradigms such as Network Slicing mandates a thorough revision of the network design with respect to the legacy approach
- Sliced networks set up a number of different network instances to run on the same infrastructure
- This makes the network management a much more complex task:
  - Resources shall dynamically be assigned to different network services
  - Their possible different QoS requirements have to be monitored in real time
- Traditionally such tasks were heavily human based, with manual configuration of the different network elements.
- This traditional way of closed loop management is not feasible anymore with novel 5G networks and beyond

Achieving closed loop automation through AI

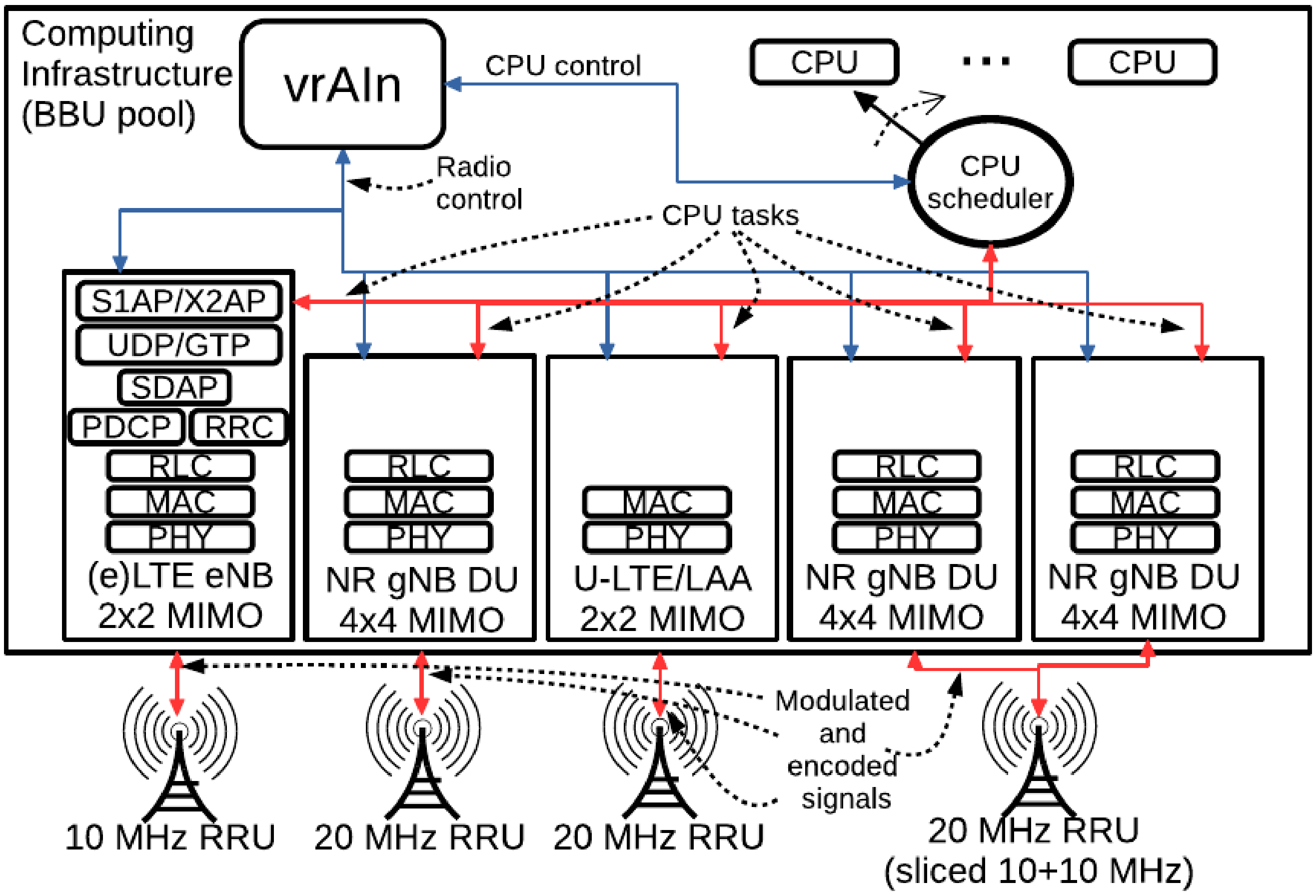
- A 5G and beyond network service management system shall
- Take advantage of the **large volume of data** flowing through the network and carrying information potentially relevant to a knowledgeable resource allocation
- Be **proactive**, by forecasting and exploiting the upcoming behaviour of a system involving many different players
- All the aforementioned tasks are among the characteristics of **Artificial Intelligence**:
- **Supervised learning** solutions can be used to perform forecasts when sufficient ground truth data can be gathered from the network
- **Unsupervised learning** solutions are fundamental when the complexity of the problem is unsuitable for traditional approaches
- **Reinforcement learning** tools are very well suited when subsequent actions are taken to maximize a certain reward

AI for network management in action:

vrAIn
A Deep Learning Approach to Virtualized Radio Access Networks (vRAN)

Why vRAN?

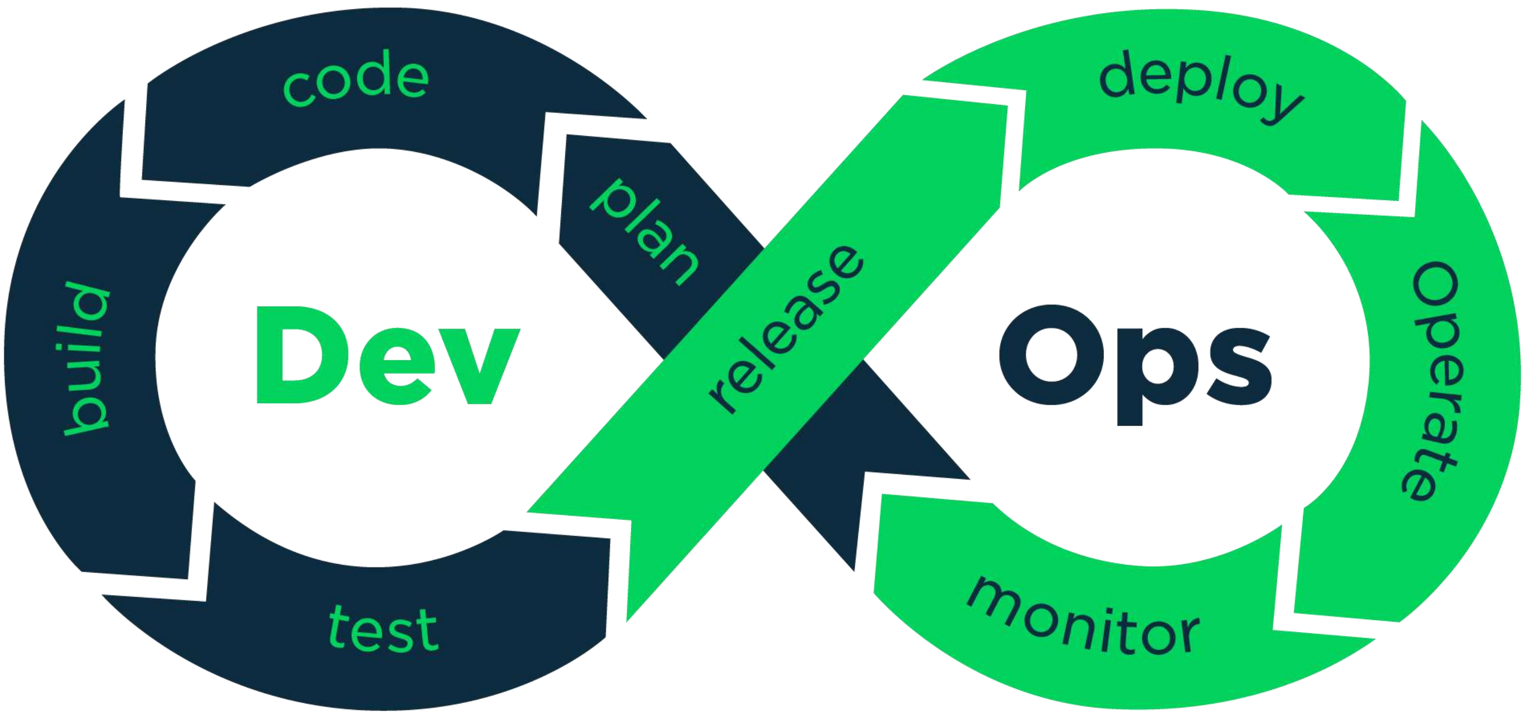
- Virtualized RAN (vRAN) **centralizes softwarized** radio access points (RAPs[1]) into **commodity** general-purpose computing infrastructure.
  - **Advantage 1:** Statistical multiplexing gains from resource pooling (via **centralization**)



[1]Base Transceiver Station (BTS) in 2G, NodeB in 3G, enhanced NodeB (eNB) in 4G, next-generation NodeB (gNB) in 5G, etc.

Why vRAN?

- Virtualized RAN (vRAN) **centralizes softwarized** radio access points (RAPs[1]) into **commodity** general-purpose computing infrastructure.
  - **Advantage 1:** Statistical multiplexing gains from resource pooling (via **centralization**)
  - **Advantage 2:** Agile update roll-ups such us security patches, protocol upgrades, bug fixes, CI/CD, agile, DevOps, etc. (via **softwarization**)



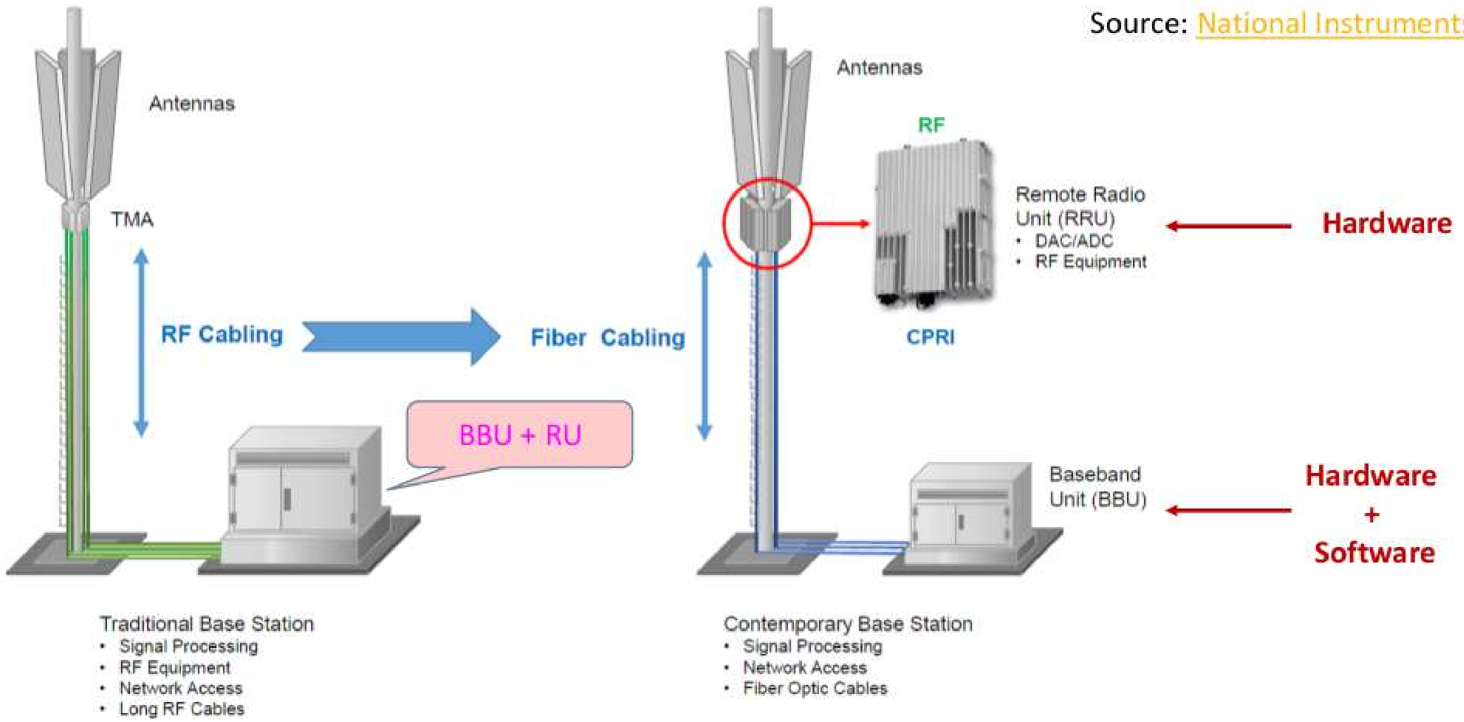
[1]Base Transceiver Station (BTS) in 2G, NodeB in 3G, enhanced NodeB (eNB) in 4G, next-generation NodeB (gNB) in 5G, etc.

Why vRAN?

Virtualized RAN (vRAN) **centralizes softwarized** radio access points (RAPs[1]) into **commodity** general-purpose computing infrastructure.

- **Advantage 1:** Statistical multiplexing gains from resource pooling (via **centralization**)
- **Advantage 2:** Agile update roll-ups such us security patches, protocol upgrades, bug fixes, CI/CD, agile, DevOps, etc. (via **softwarization**)
- **Advantage 3:** Cheaper deployment and new business opportunities (via **commoditization**)

Contemporary RAN



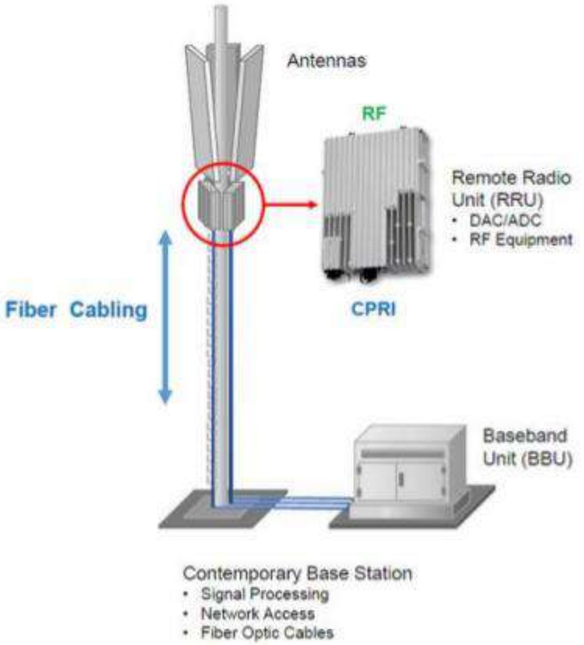
*From 3G4G Blog (http://www.3g4g.co.uk/)

[1]Base Transceiver Station (BTS) in 2G, NodeB in 3G, enhanced NodeB (eNB) in 4G, next-generation NodeB (gNB) in 5G, etc.

Why vRAN?

- Virtualized RAN (vRAN) **centralizes softwarized** radio access points (RAPs[1]) into **commodity** general-purpose computing infrastructure.
  - **Advantage 1:** Statistical multiplexing gains from resource pooling (via **centralization**)
  - **Advantage 2:** Agile update roll-ups such us security patches, protocol upgrades, bug fixes, CI/CD, agile, DevOps, etc. (via **softwarization**)
  - **Advantage 3:** Cheaper deployment and new business opportunities (via **commoditization**)

vRAN (Open RAN)



*From 3G4G Blog (http://www.3g4g.co.uk/)

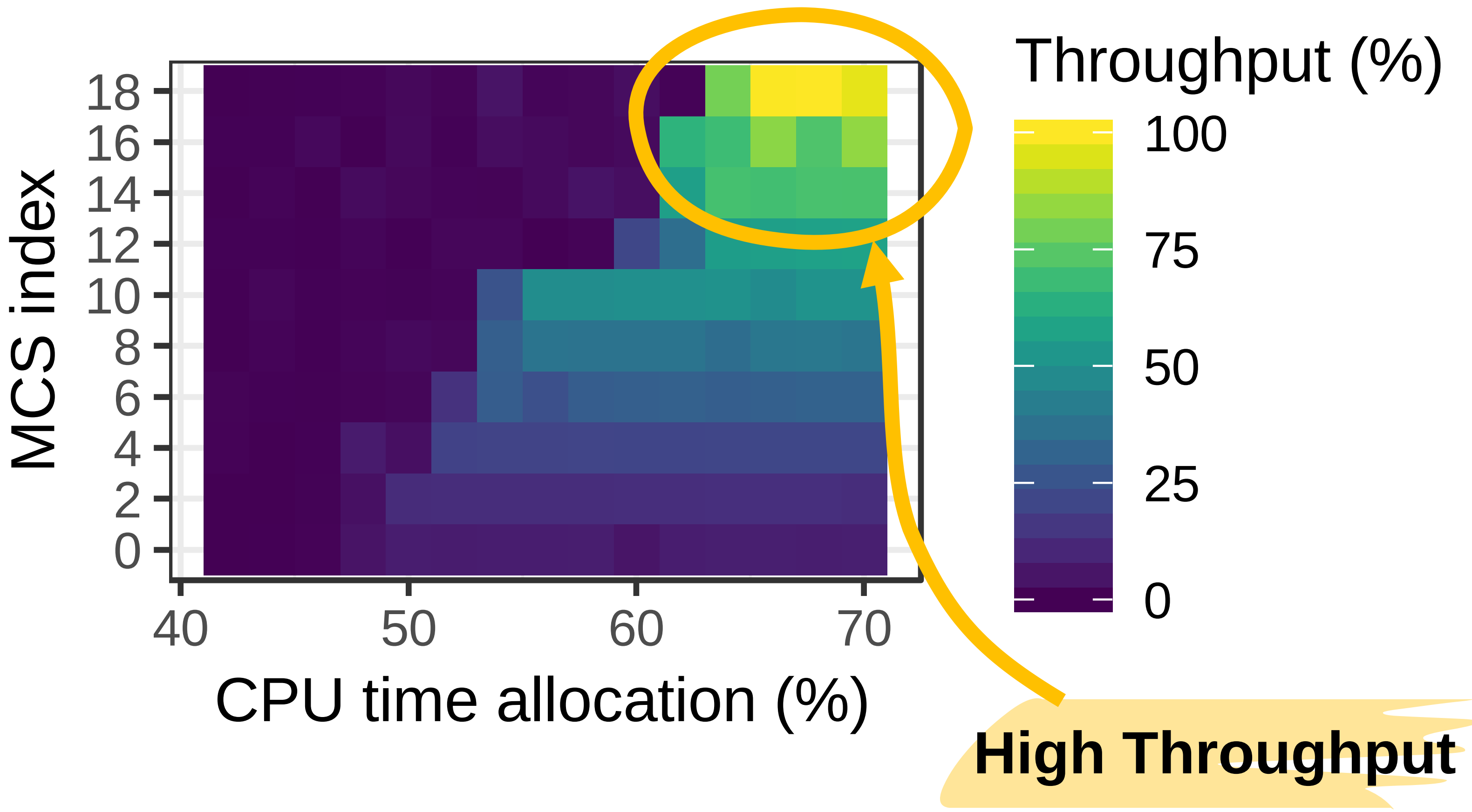
- Propietary Hardware
- GPP based COTS Hardware (SDR)
Can be purchased from any ODM/OEM/RAN hardware vendor

Open Interface
Any vendor software can work on this hardware

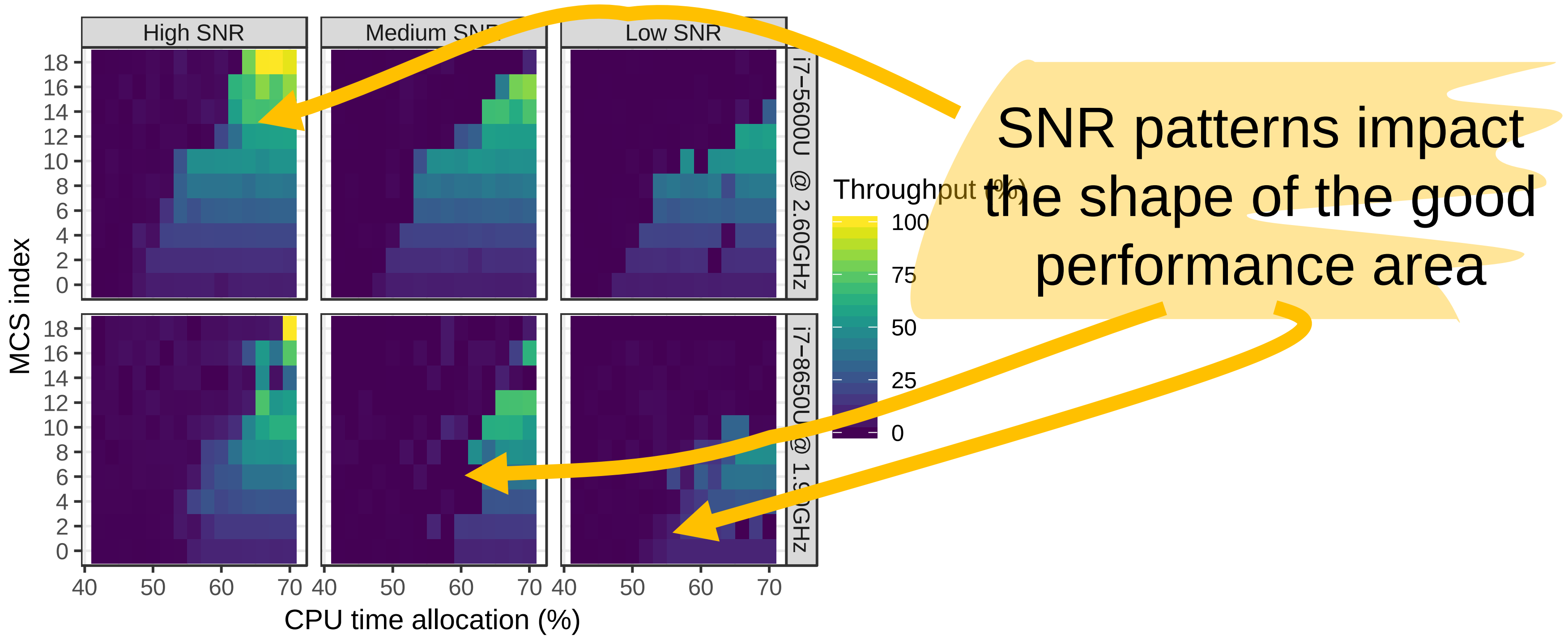
COTS Server
+
Proprietary Software with virtualized functions

[1]Base Transceiver Station (BTS) in 2G, NodeB in 3G, enhanced NodeB (eNB) in 4G, next-generation NodeB (gNB) in 5G, etc.

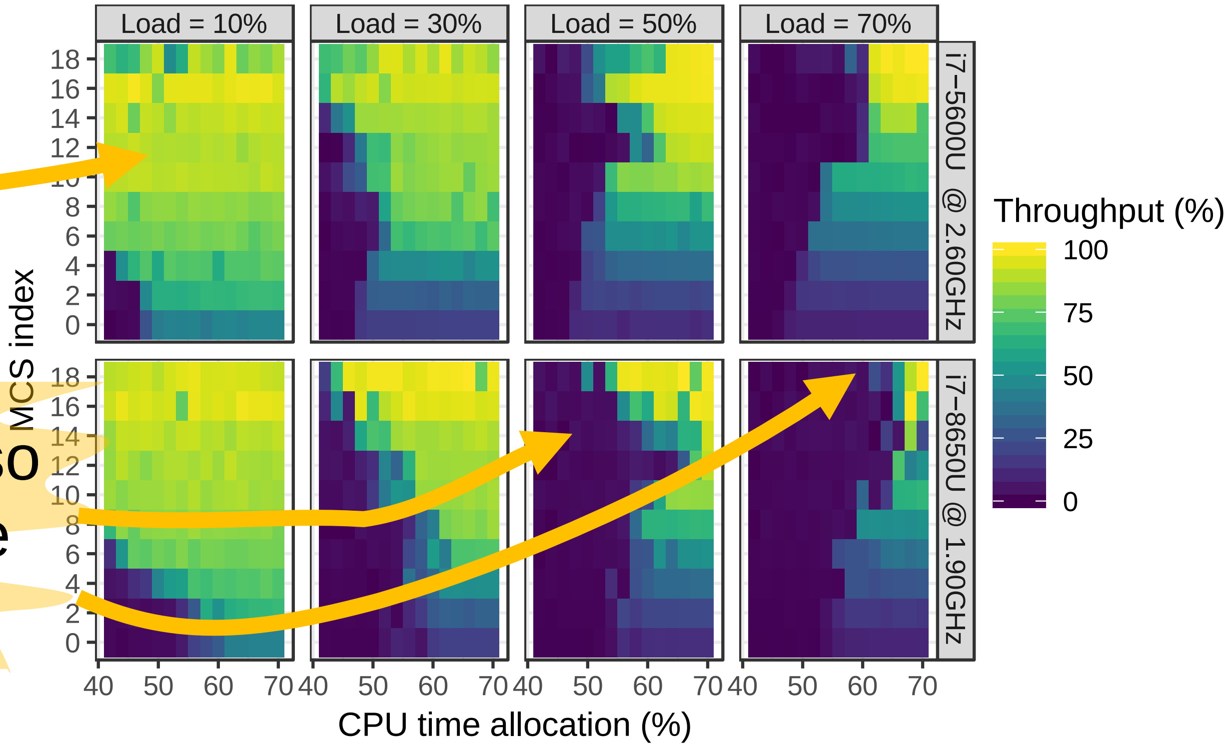
The resource orchestration problem



The problem is far from trivial



The problem is far from trivial

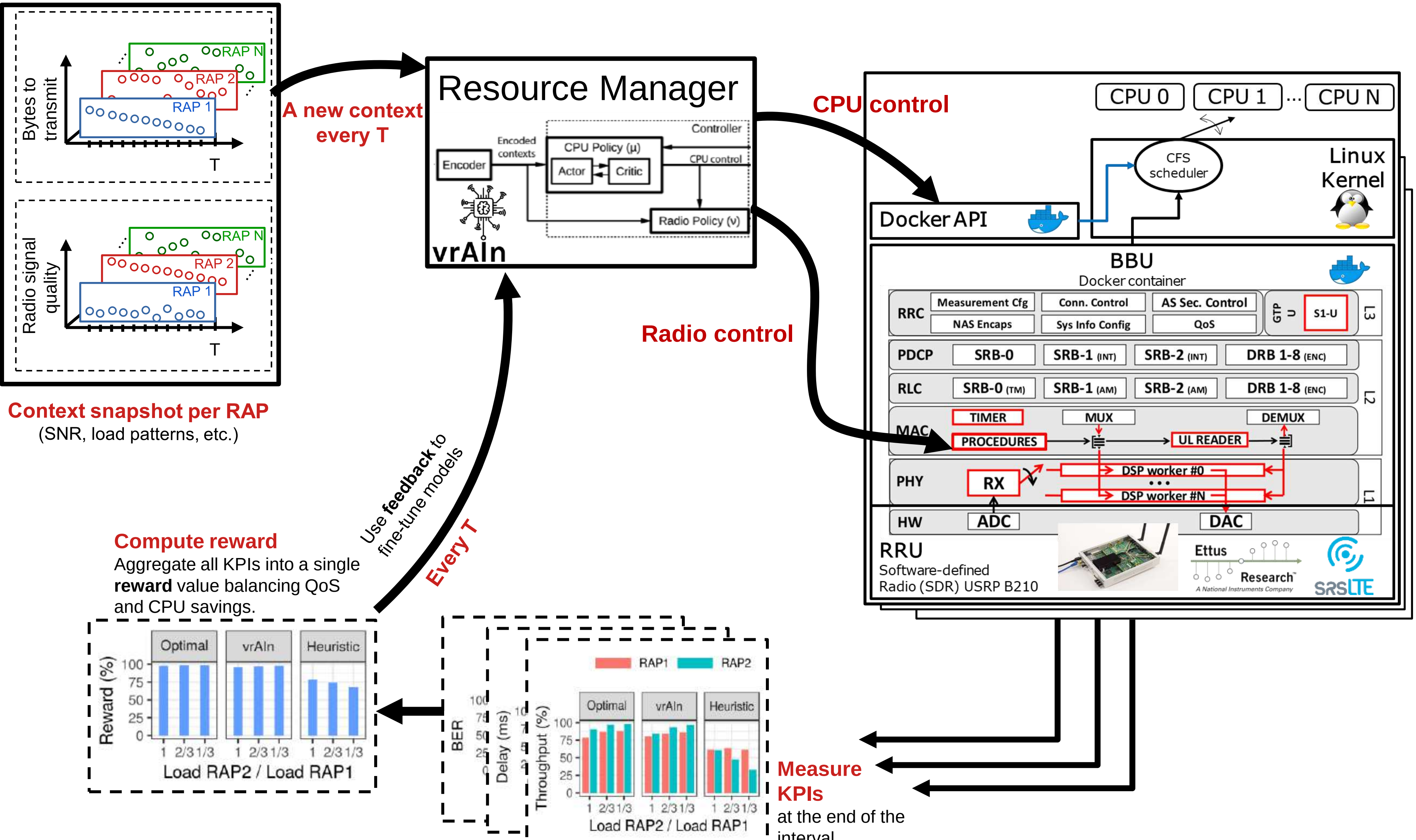


Performance also depends on the traffic patterns

The problem is far from trivial

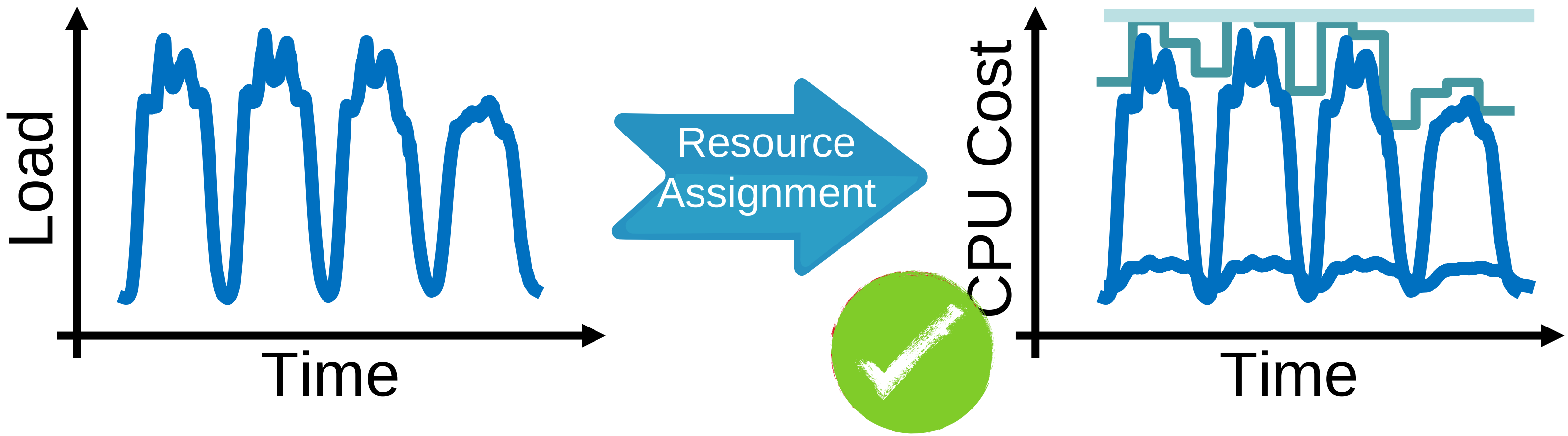
Performance is a very complex function of the contexts and the resource assignment
→ Deep Learning

vrAIn: AI based vRAN resource controller



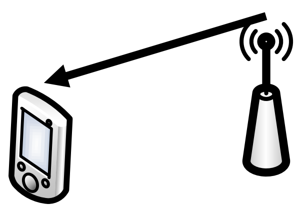
Backup

The resource orchestration problem

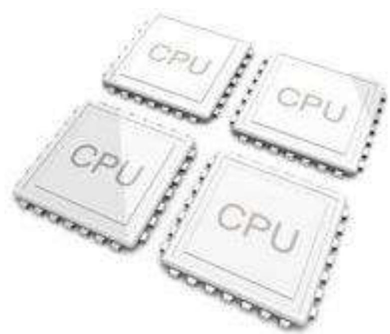


Resource assignment depends on many factors such as…

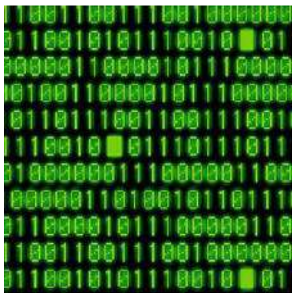
User Demand



Channel
Conditions



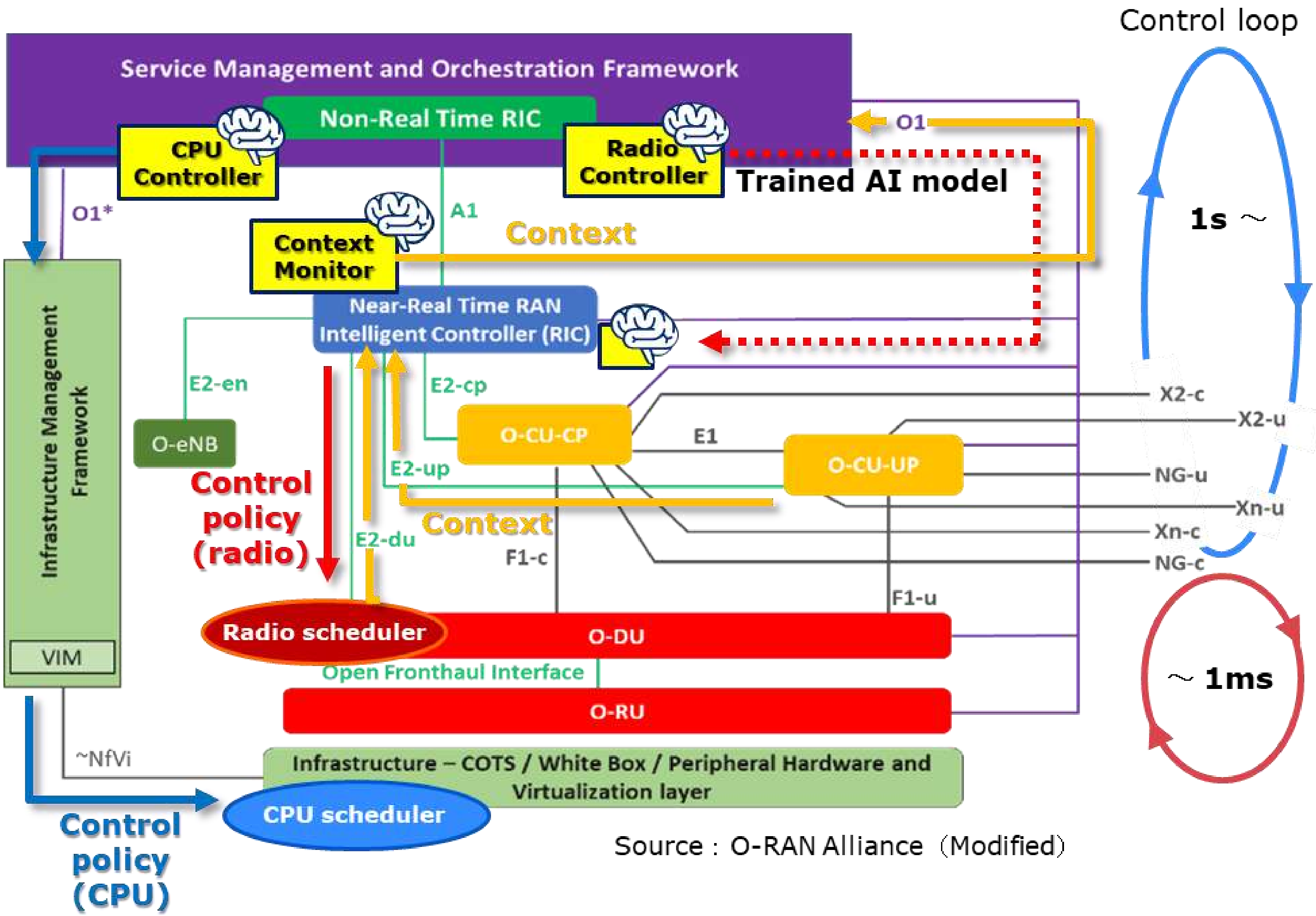
CPU
Platform



SW stack,
functional split

Integration of vrAIn into O-RAN

vrAIn



Source： O-RAN Alliance（Modified）

Reward function

$$r(\mathbf{x}, \mathbf{a}) := \sum_{i \in \mathcal{P}} \mathbb{P}\left[q_{i,x_i,a_i} < Q_i\right] - M\varepsilon_i - \lambda c_i$$

Decoding Error Probability (ε_i)
Empirically computed by sampling every subframe (UL) or via HARQ

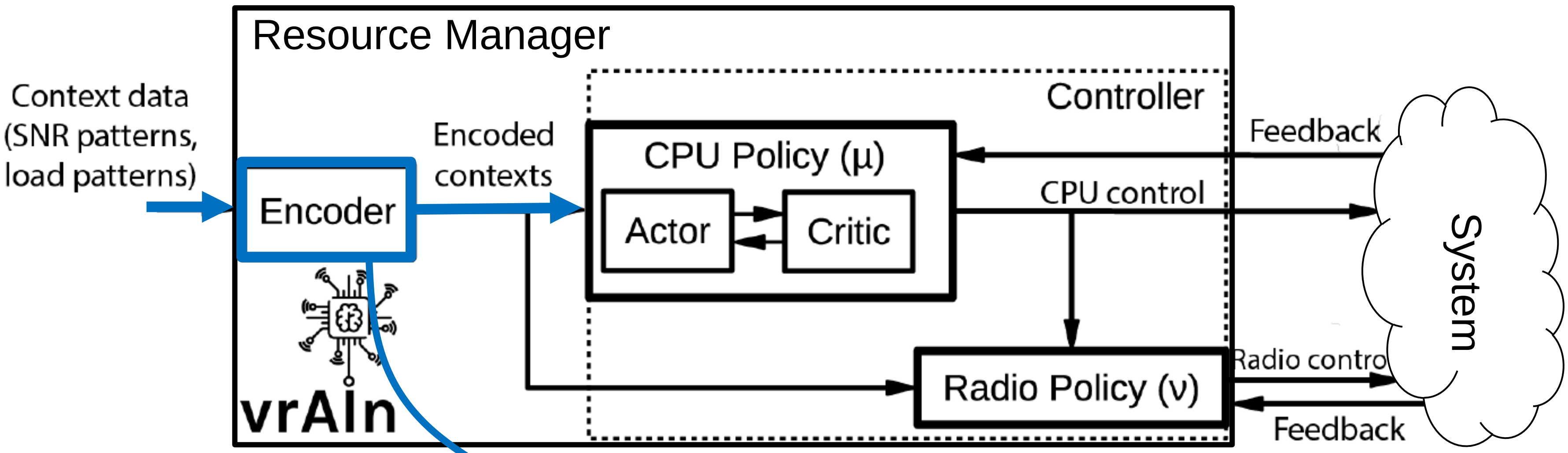
CPU allocation (c_i)

Buffer State (random variable) (q_{i,x_i,a_i})
Empirically estimated by sampling every Buffer State Report (BSR) in UL

Target mean buffer state (bytes) (Q_i)
An easy measure of delay

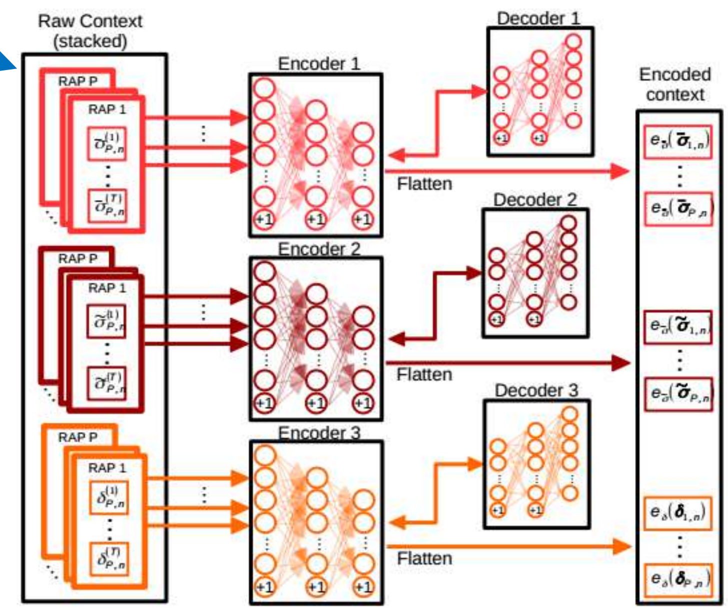
Constant Parameters (M, λ)
M is a large value
λ is a small value

vrAIn: Challenges and Solutions

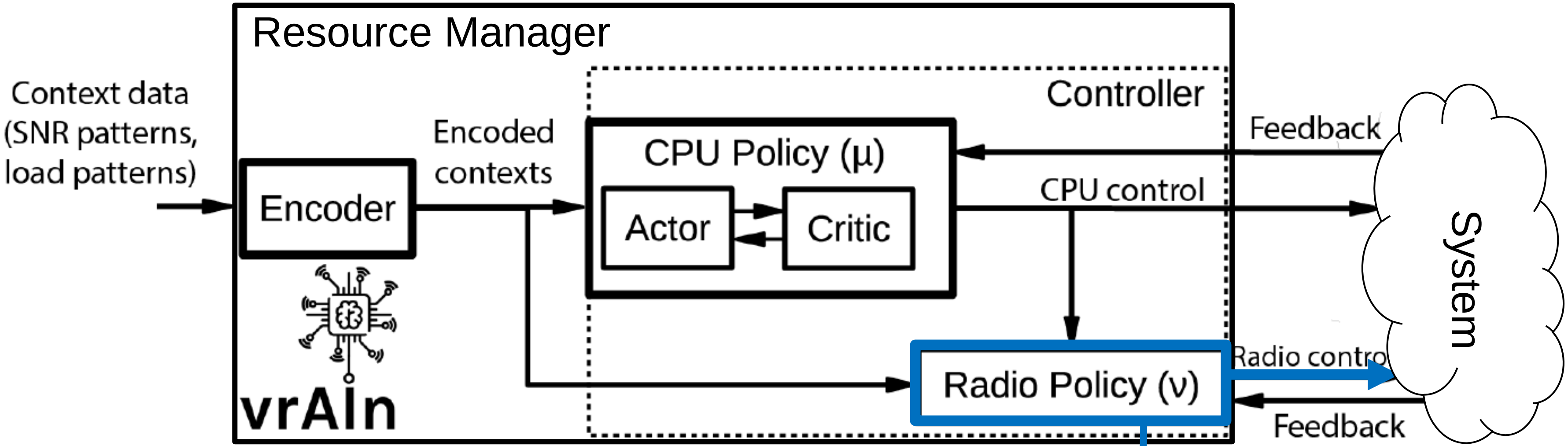


Challenge #1:
Dimensionality of the input contexts

Solution:
3 Sparse autoencoders to reduce the dimensionality

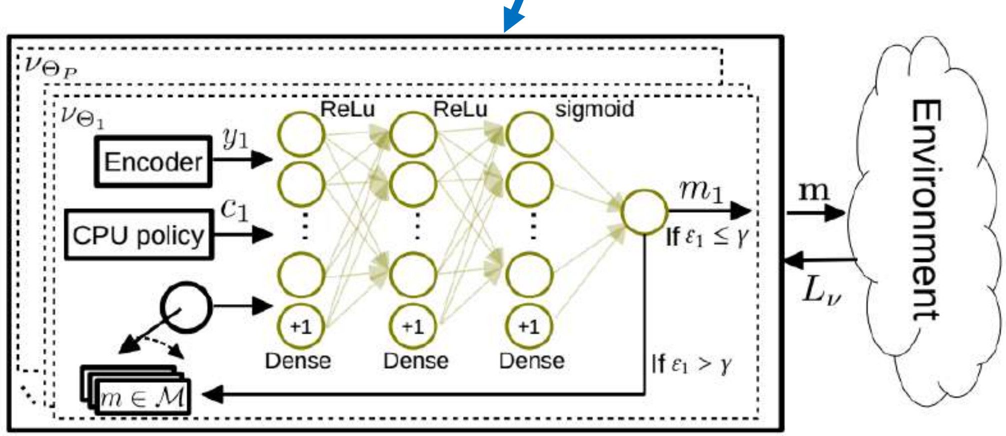


vrAIn: Challenges and Solutions

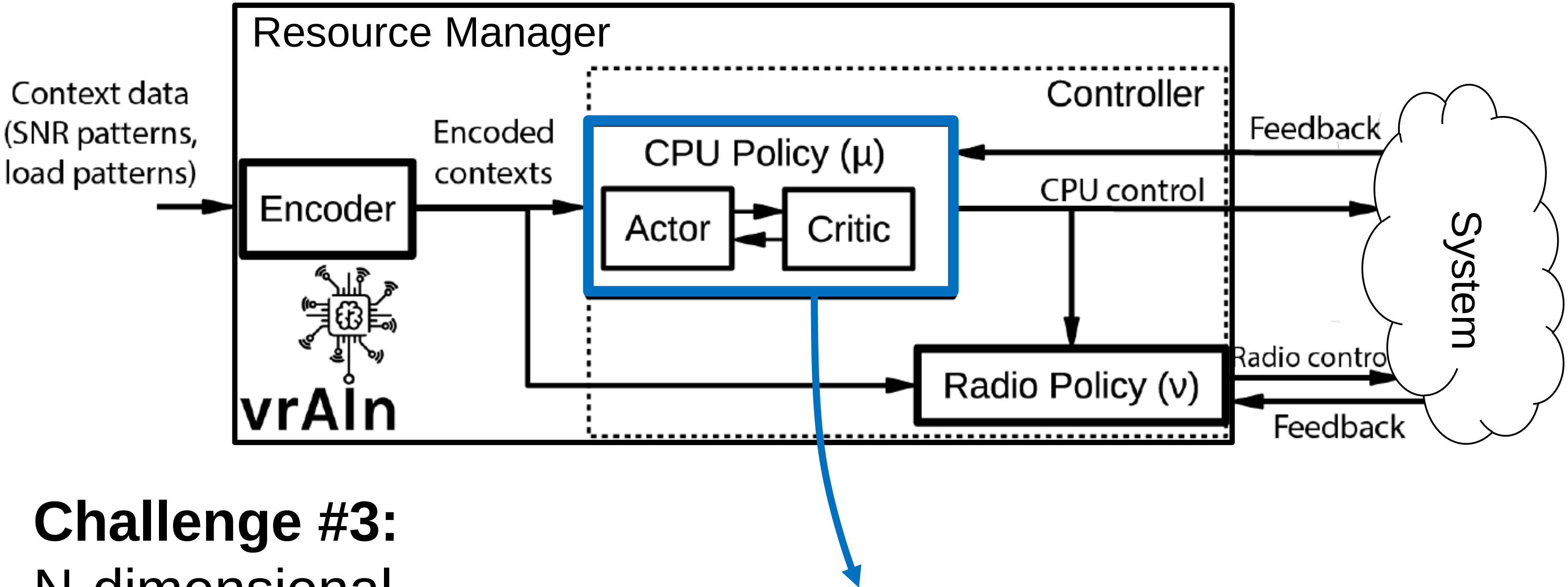


Challenge #2:
Heterogeneity of the action space (continuous and discrete)

Solution:
Decoupling of the radio and the CPU policy

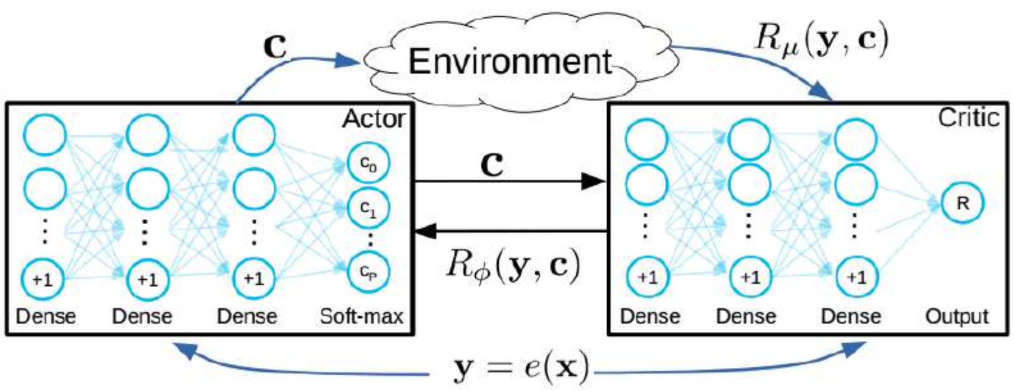


vrAIn: Challenges and Solutions



Challenge #3:
N-dimensional continuous controls for the CPU policy

Solution:
Deep Deterministic Policy Gradient



Evaluation results: Unlimited Resources

Scenario 1

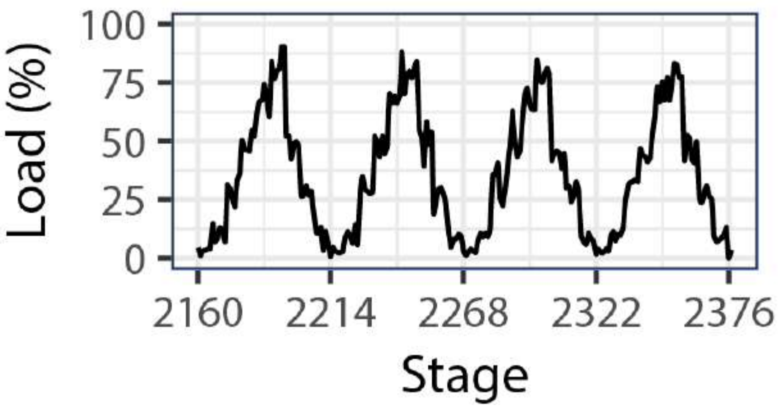
- Unlimited CPU resources
- One virtual Base Station

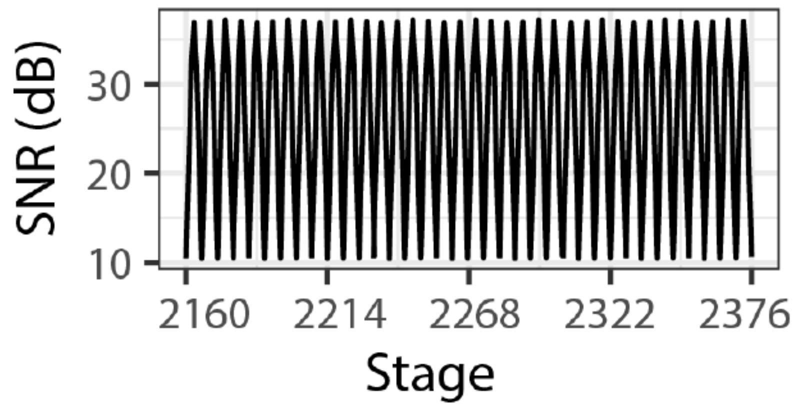
Objective:

- Minimize the costs while satisfying the QoS

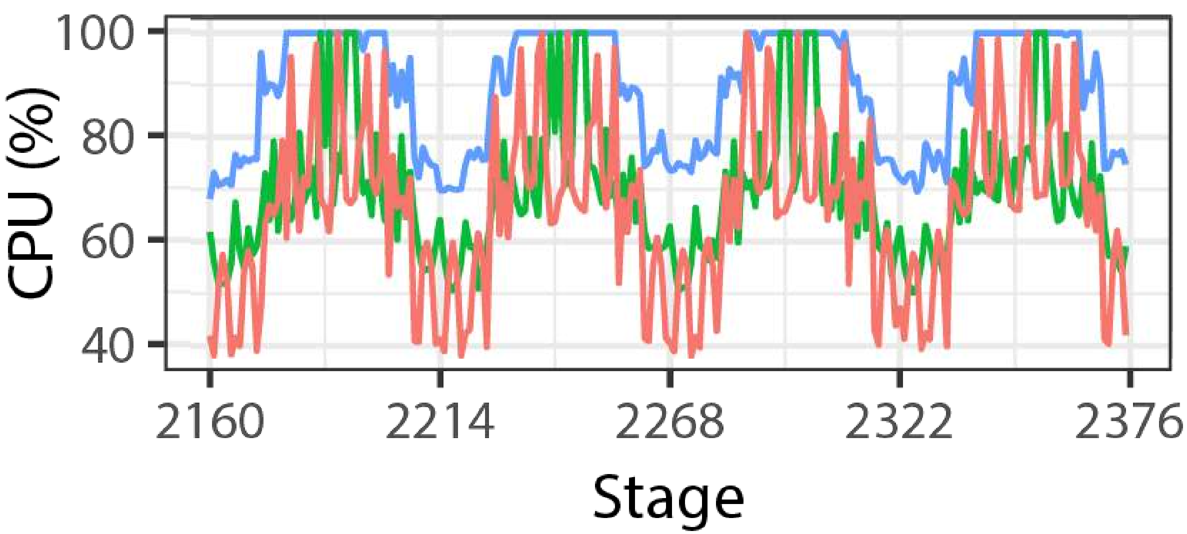
Evaluation results: Unlimited Resources

Contexts:





vrAIn CPU allocation:



	Average CPU Savings
High QoS	14%
Medium QoS	26%
Low QoS	39%

Evaluation results: Limited Resources

Scenario 2

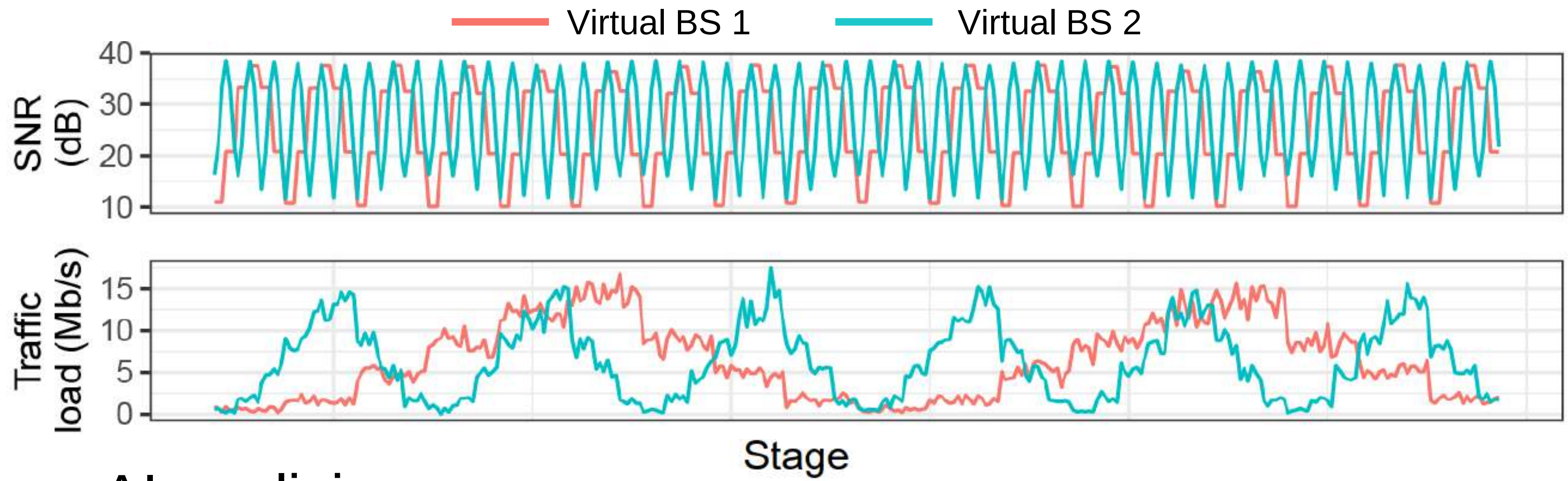
- Limited CPU resources (one core)
- Two virtual Base Station

Objective:

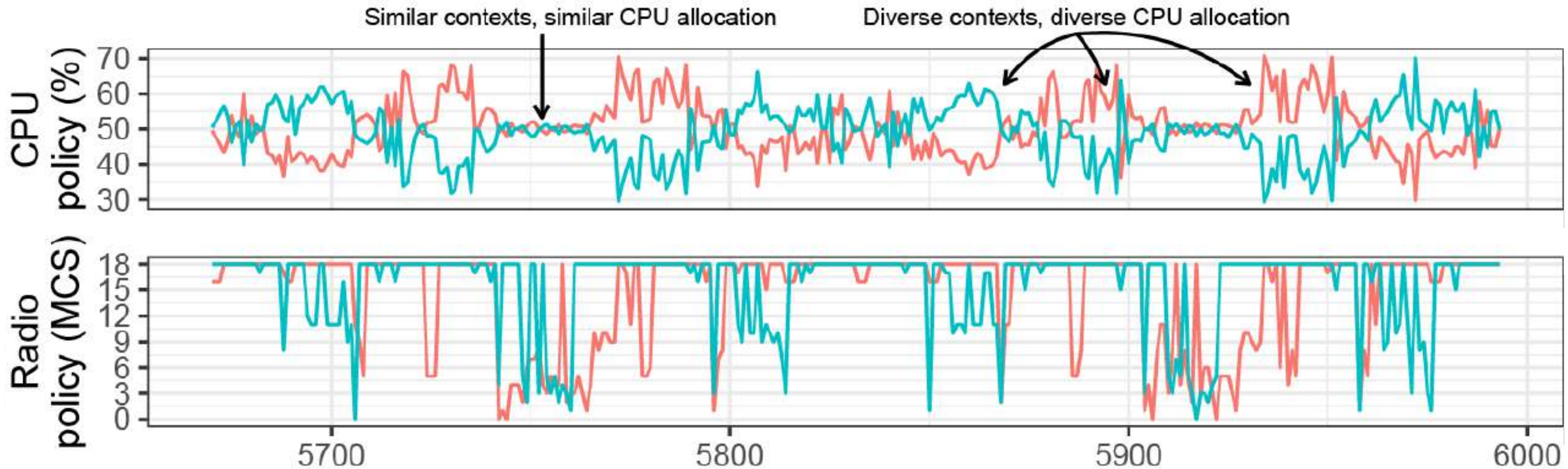
  - Maximize the performance of both virtual BSs

Evaluation results: Limited Resources

Contexts:



vrAIn policies:



vrAIn achieves zero decoding error

Conclusions

- The performance of a virtual BS is a very complex function of the contexts and the resource assignment, motivating the use of **Deep Learning**.

- We solve the problem using a novel combination of **Sparse Autoencoders**, a **Reinforcement Learning** algorithm and a **Neural Network Classifier**.

- Our solution **minimizes the costs** with unlimited resources and **maximizes the performance** with limited resources. With respect to state-of-the-art solutions, vrAIn achieves…
    - **CPU savings ~30% with unlimited resources.**
    - **Throughput increase ~25% per virtual Base Station.**

- We trained our models with **real data** and implemented a **proof-of-concept** of the solution.
    - **Dataset in https://github.com/agsaaved/vrain**