Technical Workshop:
Academia

Ariel G. De Castro
Alan Teixeira
Fabrício Rodríguez
Francisco G. Vogt
Christian E. Rothenberg

Schedule

Selected research ideas candidate for experimental evaluation in OpenRAN Brasil

- Federated learning for INT collection
- Synthetic packet trace generation (PCAP)
- P4 Testbed evolution

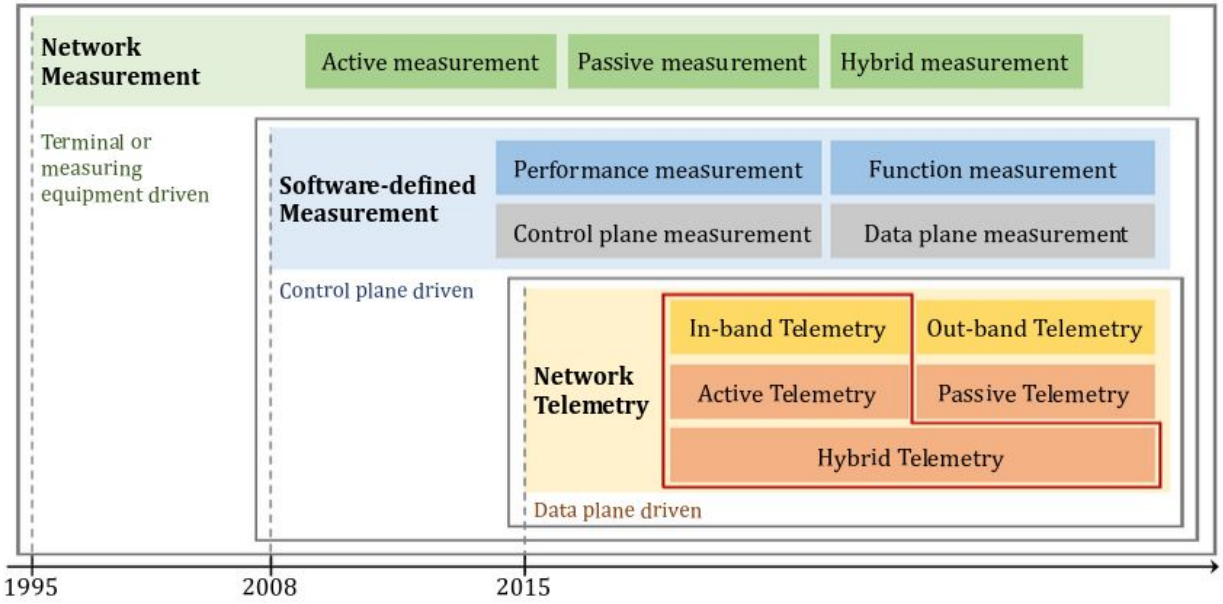
UNICAMP

Federated Learning for INT collection

openRAN BRASIL

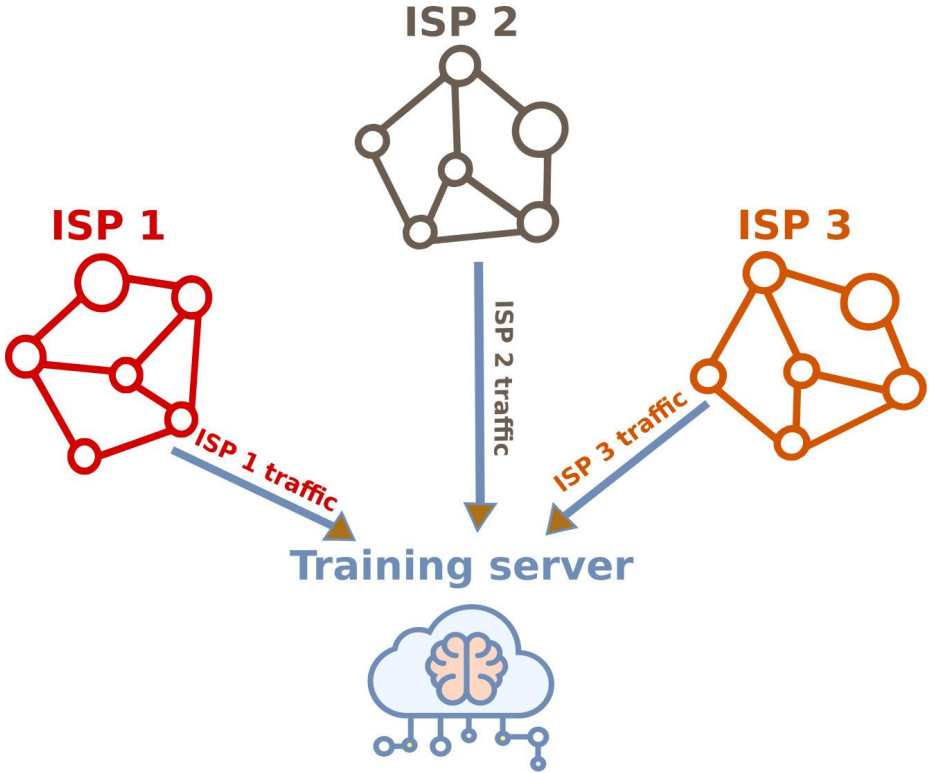
Introduction: Motivation

With P4 and INT, we have seen increasing efforts towards autonomous network decisions

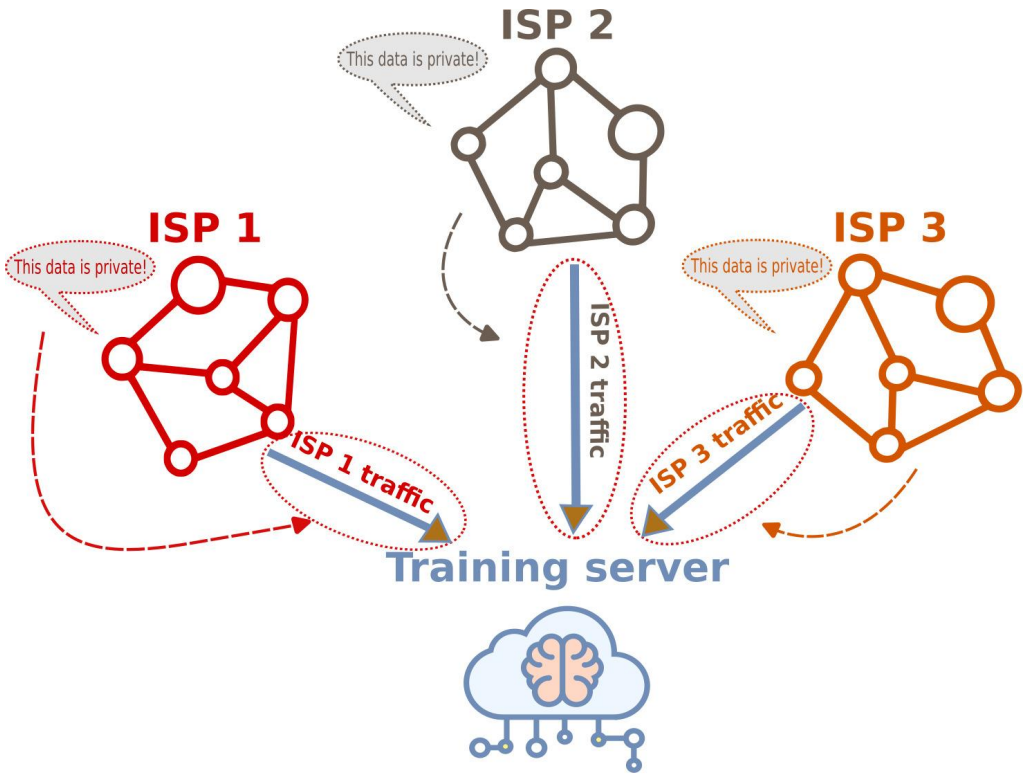


Source: [Tan et al., 2019 - In-band Network Telemetry: A Survey]

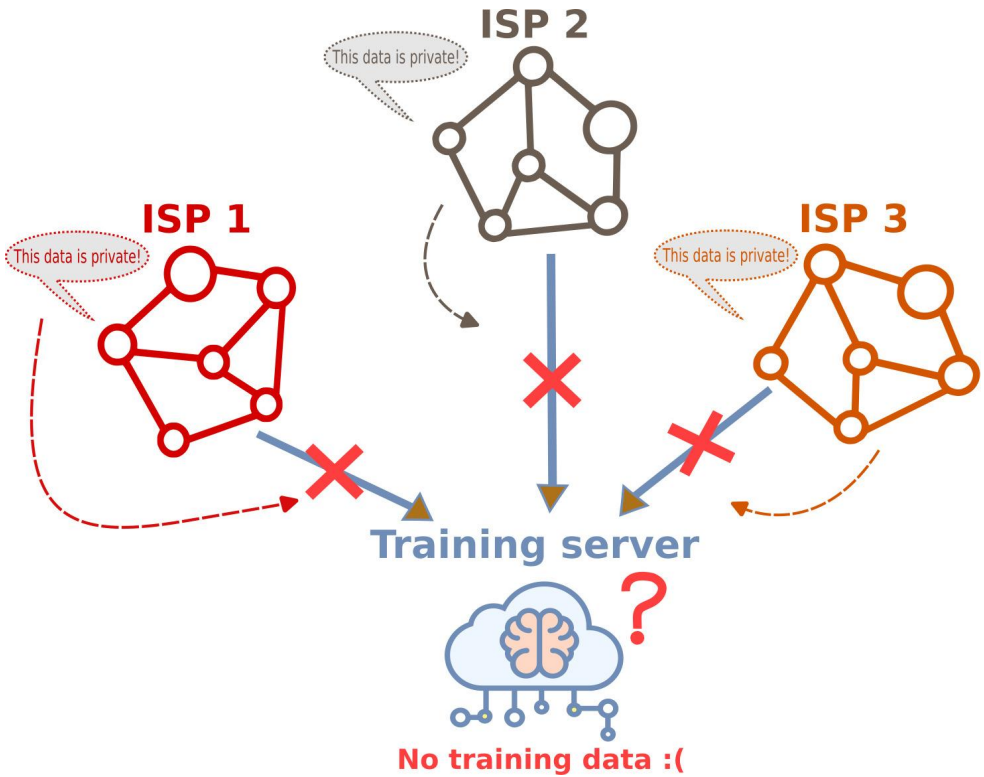
Introduction: Problem (1/3)



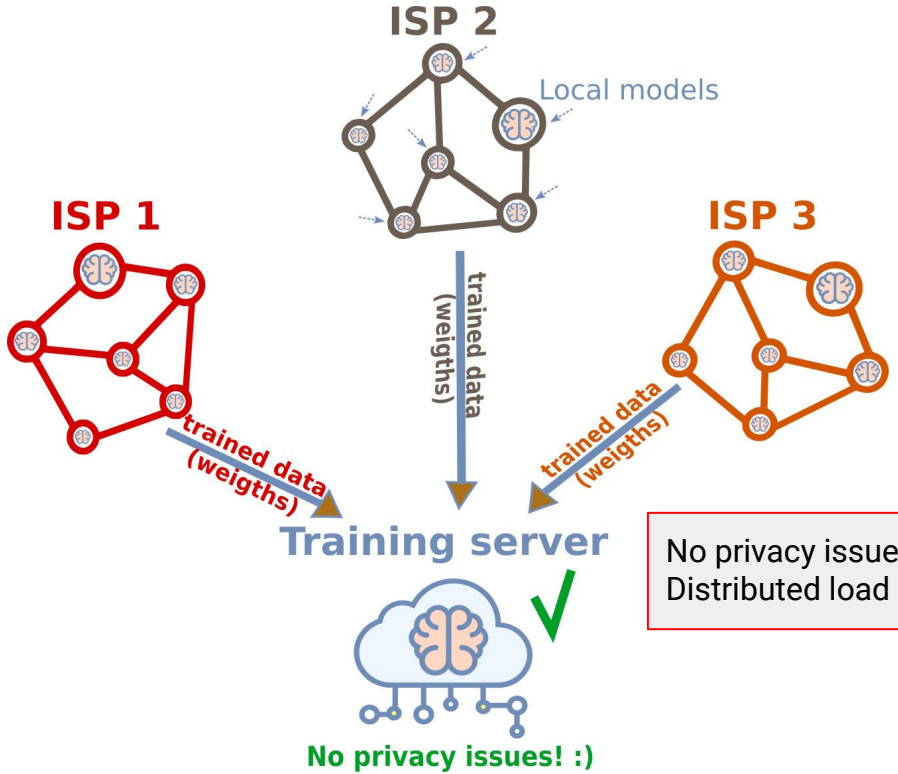
Introduction: Problem (2/3)



Introduction: Problem (3/3)



Introduction: Architecture



No privacy issues!
Distributed load across the edges!

Related work (Traffic Generation w/ GAN)

Authors	Proposal
Wang et al., 2019	Disseminates ML models across multiple edges to save money
Wang et al., 2020	Training of ML models distributed across fog nodes
Jin et al., 2023	Online algorithm for INT collection. Generic model for federated learning
Ji et al., 2023	INTService: an online system that processes INT queries and utilizes a primal-dual mechanism to reduce energy consumption

Status: Federated training

Pros:

- Privacy: little exposure to private data (e.g., user data, vulnerability exposure)
- Automated decisions (i.e., little to no operator's action needed)
- Less energy consumption (i.e., OPEX), since the training load is distributed

Cons:

- Global model and local models must be synchronized
- Each local model must be synchronized and updated to the global model
- The training frequency must not be **too** bandwidth consuming

UNICAMP

Synthetic packet trace generation (PCAP)

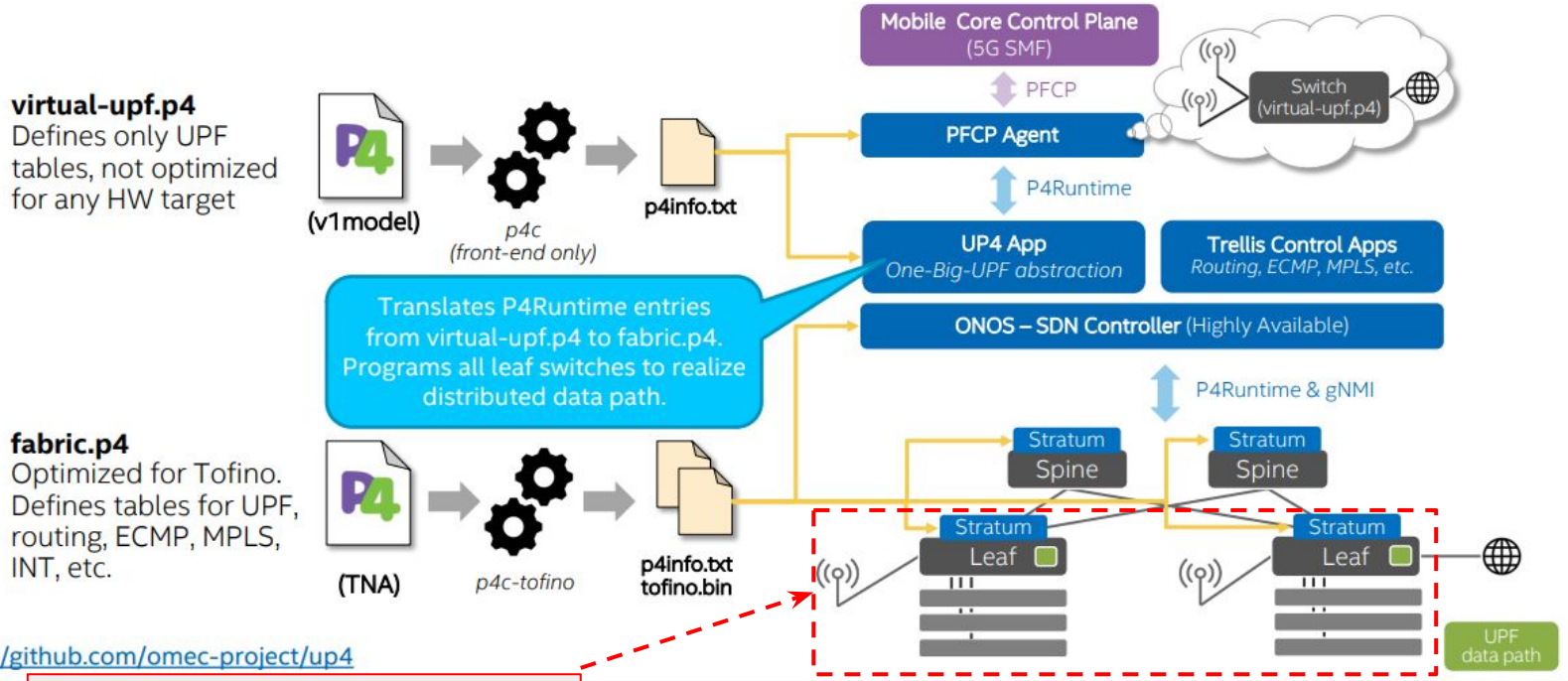
Realistic Traffic Generation

openRAN BRASIL

Recap: P7

openRAN BRASIL

P7 + P4 (1/2)



We can leverage P7 to emulate UPF functionalities and link characteristics in different scenarios

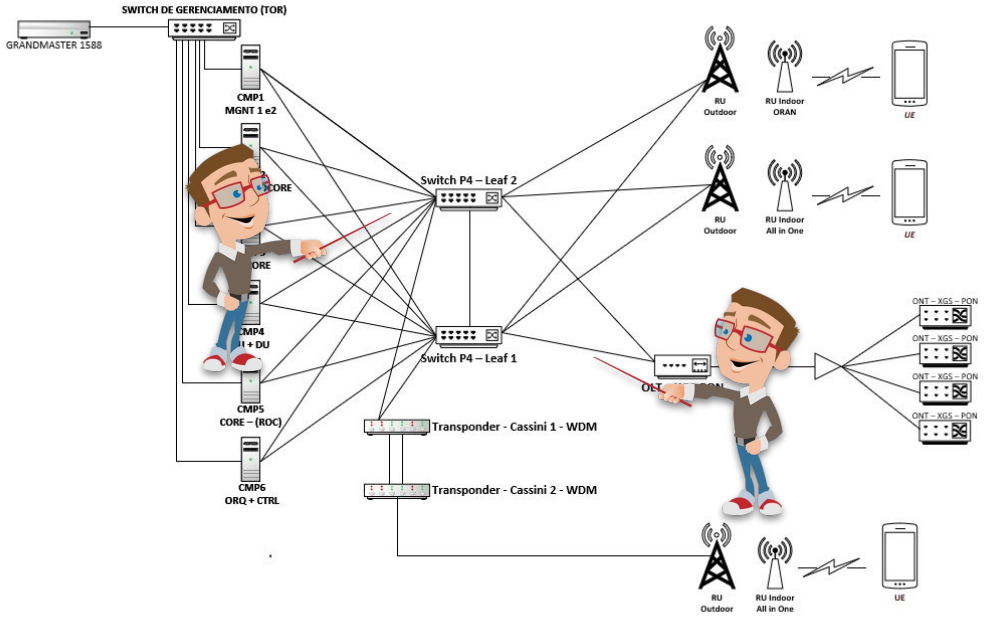
P7 + P4 (2/2)

Physical Testbed architectural options

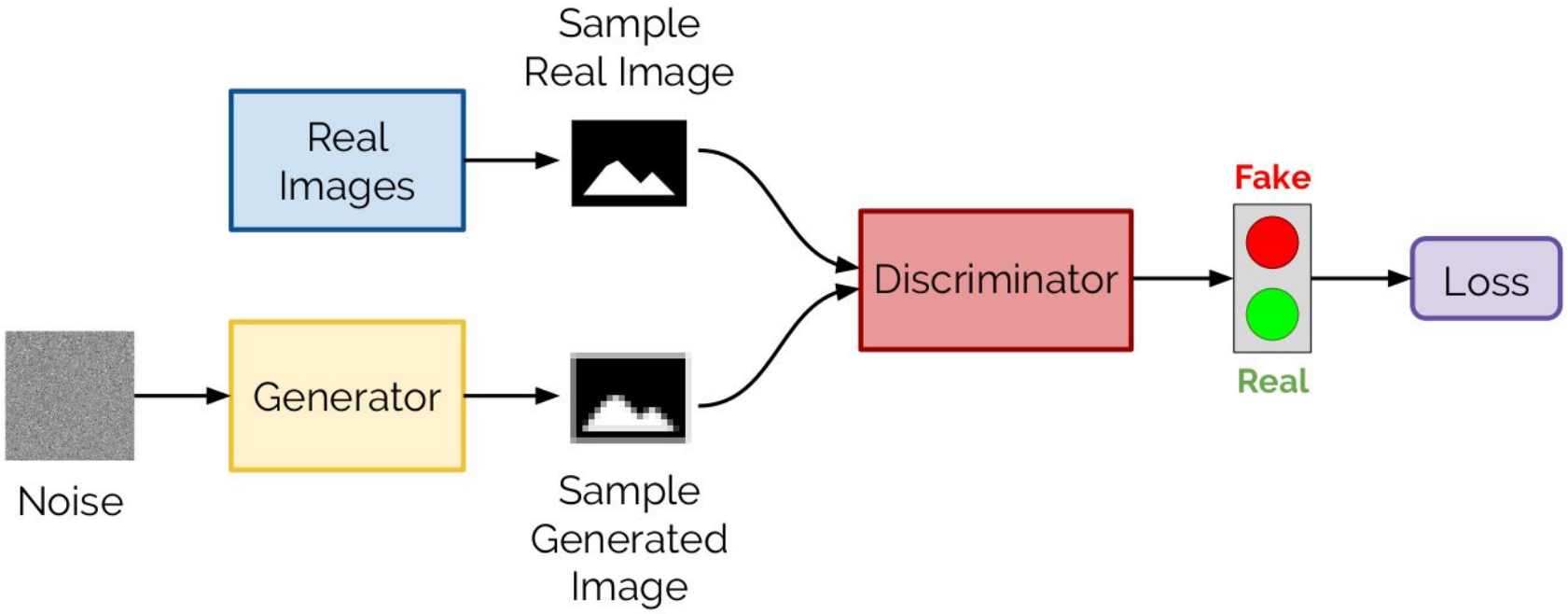
- Bump-in-the-wire P7 emulation
  - Allocate one physical Tofino switch into the topology
- Shared Tofino
  - 1 pipe for P7
  - 1 pipe for UPF

Further considerations

- INT support
- Barefoot Runtime dynamic configuration of tables and ports
- 100G workloads

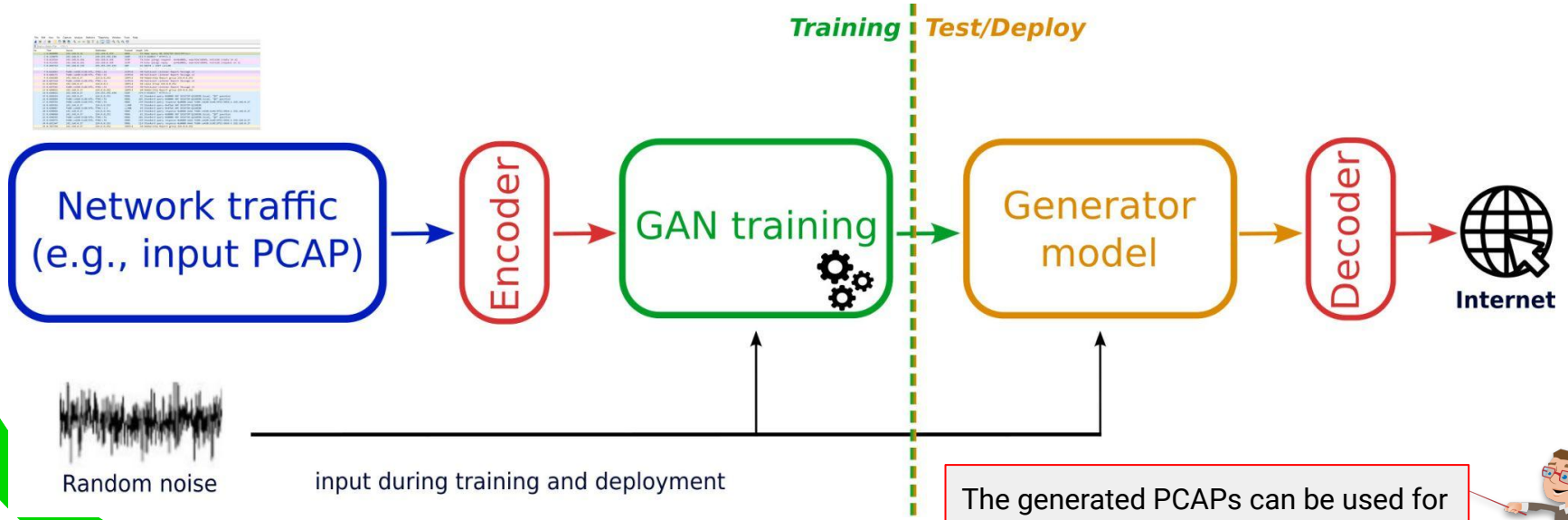


Introduction: Motivation



Source: https://cpang4.github.io/gan/

Architecture: PCAP generator



The generated PCAPs can be used for multi-purpose (e.g., penetration tests, worload, traffic behavior for heterogeneous applications)

Related work (Traffic Generation w/ GAN)

Authors	Proposal
Brock et al., 2018	A BiGAN method for packet generation
Dowoo et al., 2019	PcapGAN: focused on cyber data. It generates packet traces by considering the header order and temporal generation (e.g., TCP 3-way-handshake)
Meddahi et al., 2021	SIP-GAN: an extension of GANs model for SIP, aiming to process and generate SIP traffic at packet level
Anande et al., 2023	Two GAN architectures that include data transforms to simultaneously train and generate categorical and continuous network traffic features

Status

Under discussion:

- PyTorch or Tensorflow?
  - Tensorflow = production [?]; Pytorch = research[?]
- CPU only processing? GPU too?

Done:

- We already have access to CAIDA datasets

Pros:

- Generate as many data as needed, since real packet traces are private
- Generate realistic data for several use-cases (e.g., threats analysis, QoS)

Cons:

- Hard to implement a generic model for multiple protocols simultaneously (i.e., overfitting)

Final remarks

- Federated Learning has the potential to enhance the automation of network decisions by increasing their reliability.
    - Distributed training process has the potential to solve the problem of privacy in multiple ISPs.
- PCAP generation enables precise application data behavior analysis.
- These ideas are in their initial stages, not in any well-defined.

UNICAMP

Thank you!

INTRIG INFORMATION & NETWORKING TECHNOLOGIES RESEARCH & INNOVATION GROUP

openRAN BRASIL