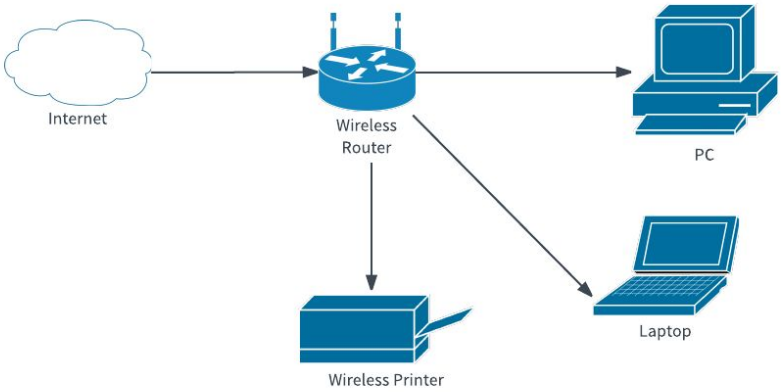
PINOT: Programmable Infrastructure for Networking

Roman Beltiukov, Sanjay Chandrasekaran, Arpit Gupta, Walter Willinger

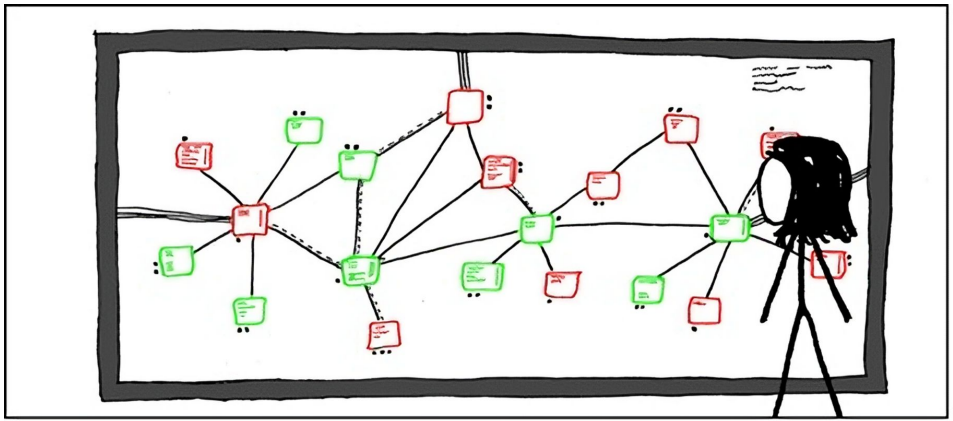
https://pinot.cs.ucsb.edu

(One of) academia's problem: representative infrastructures

Typical



Desired



Results: **bad** data ➔ **bad** solutions

(e.g., see "AI/ML for Network Security", https://doi.org/10.1145/3548606.3560609)

Example platforms

- RIPE Atlas https://atlas.ripe.net
- COSMOS https://www.cosmos-lab.org
- Netrics https://github.com/internet-equity
- Measurement Lab https://www.measurementlab.net/tests
- Cloudlab https://www.cloudlab.us/
- Iris https://github.com/dioptra-io
- PerfSonar https://www.perfsonar.net
- EdgeNet https://www.edge-net.org/
- OneLab https://onelab.eu/
- EmuLab https://www.emulab.net/
- ...

They are great, but…

Some experiment are hard (or impossible) to implement, e.g.:

- YouTube Quality of Experience
- + Over Wi-Fi
- + In a live network with users
- + Over long time period
- + Flexible and programmable client
- + Separate backbone problems from last-mile problems

What infrastructure do we need for this?

Our solution: programmable infrastructure @ UCSB campus

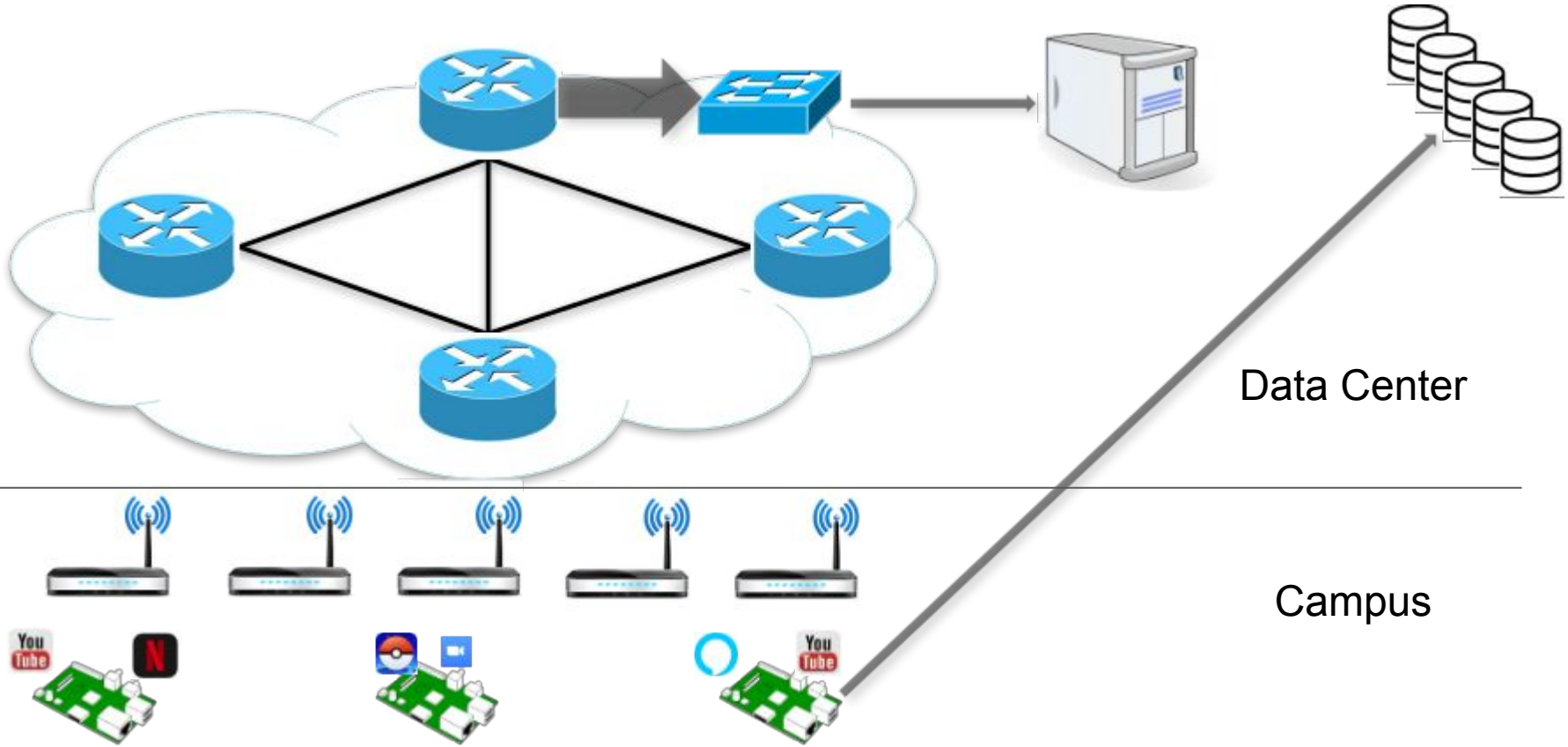
Design principles

- Active + passive measurements
- Last-mile connection carrying real-world user traffic
  - Balance between “unrealistic lab scenario” and “inaccessible production network”
  - Mimics a typical enterprise network, 25k+ users
  - Latency spikes, sudden user & traffic overloads, peak hours, etc
- Localized deployment (nodes are close geographically and logically)
- Support for **arbitrary** experiments (Docker-based)
- Direct and fast access for research iterations (fail-fast)

Additionally:

- Ethical: minimal disruption + preserve privacy
- Fully reproducible: cheap components + everything open sourced

Overall architecture



Active measurements

60+ Raspberry Pi 4 devices (40 more in this month)

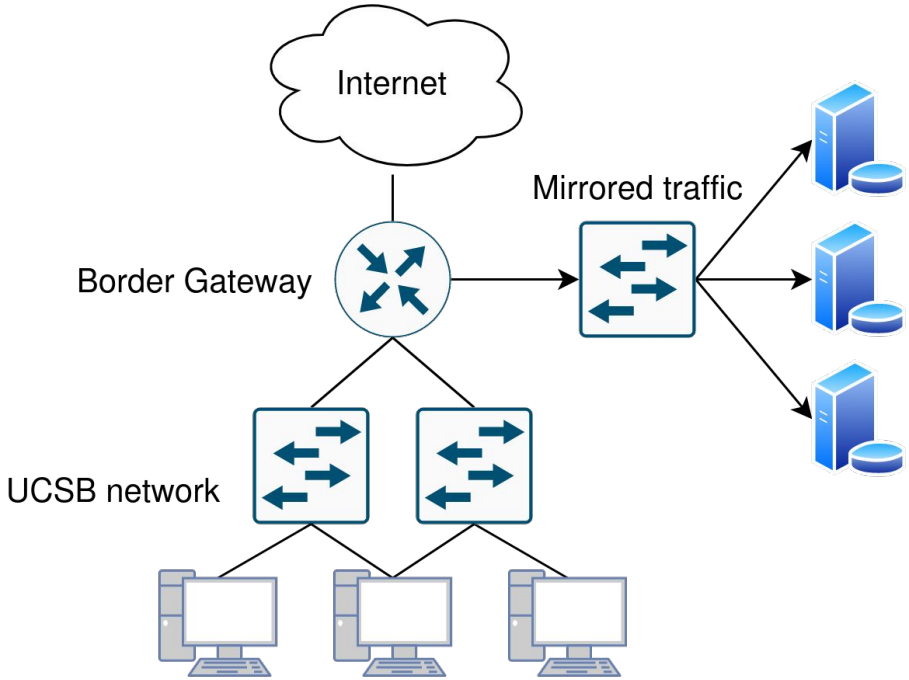
- Controlled by SaltStack
- Deployed in public places
- Use UCSB Wi-Fi infrastructure (and sometimes wired)
- Mimic real users

Raspberry Pi 4 + PoE + Ubuntu


MILLS WAY
SHADOWBROOK DR.
HIGH GROVE AVE.
Whittier
SWEETWATER WAY
Santa Catalina
EL COLEGIO ROAD
CERVANTES ROAD
EL GRECO ROAD
PICASSO ROAD
PICASSO ROAD
SEGOVIA ROAD
CORDOBA ROAD
PARDALL ROAD
ABREGO ROAD
Abrego & Camino Corto
ESTERO ROAD
66 Block
65 Block
SUENO ROAD
FORTUNA ROAD
67 Block
68 Block
PASADO ROAD
Isla Vista
CAMINO MAJORCA
CAMINO CORTO
CAMINO DEL SUR
CAMINO PESCADERO
SABADO TARDE RD.
SABADO TARDE RD.
DEL PLAYA DRIVE
DEL PLAYA DRIVE

Passive data collection

- Intel Tofino Switch for live traffic mirroring
- ONTAS (P4) anonymization on the switch
- Three servers for balancing data collection



Active + Passive measurements benefits

- Active measurements
  - Full programmable control of clients
  - **Labelled data! :)**
- Passive measurements
  - Live real-world traffic
  - Data diversity, time patterns, network events
- Active + Passive combination
  - Multiple vantage points for packet observation
  - Enrich labelled data with unlabelled (but similar) traffic

You can implement and use together or separately

Current experiments examples

- Video Quality of Experience measurements (YouTube vs Twitch vs Vimeo)
  - See motivation example
- Google Meet & Zoom QoE measurements
- Controlled speed tests (time, interface, location)
- Application traffic collection for fingerprinting
- Botnet imitation with network attacks :D

Limitations & Practical issues to be aware of

- Theoretical data guarantees are debatable
  - Solution: measure, explore, and confirm
- Ethical Review is **important** and **required**
  - Solution: do it **before** buying anything
- Official university services could be slow :)
  - Solution: start **early**
  - Solution: find students/personnel with direct access to buildings/services you need
- Chips/RPi/components shortage
  - Solution: RPi-like boards + Armbian instead of Ubuntu
- RPi OS image security is **important**
  - Solution: google “Linux hardening” and **do it**
  - Exception: LUKS is slow ><

TL;DR:

- We've created active+passive measurement platform @ UCSB
  - Pretty unique active programmable measurements + concurrent passive live traces collection
- It has live user traffic (25k+ users)
- It is cheap and reproducible
  - All hardware components are easy to buy
  - All software components are open-source or publicly available
  - See "Reproducibility" page of the website for all repos and information
- We invite other researchers to submit experiments
  - pinot@cs.ucsb.edu

https://pinot.cs.ucsb.edu