2022 IEEE International Conference on Big Data (Big Data 2022)

HAIVAN: a Holistic ML Analytics Infrastructure for a Variety of Radio Access Networks

Hong-Linh Truong
Department of Computer Science
Aalto University, Finland
https://rdsea.github.io

Nguyen Ngoc Nhu Trang
Central MobiFone Network Centre,
MobiFone Corporation, Vietnam
trang.ngocnhu@mobifone.vn

Outline

- **V-RAN**
- **Holistic ML infrastructure requirements**
- **HAIVAN methods and designs**
- **Conclusions**

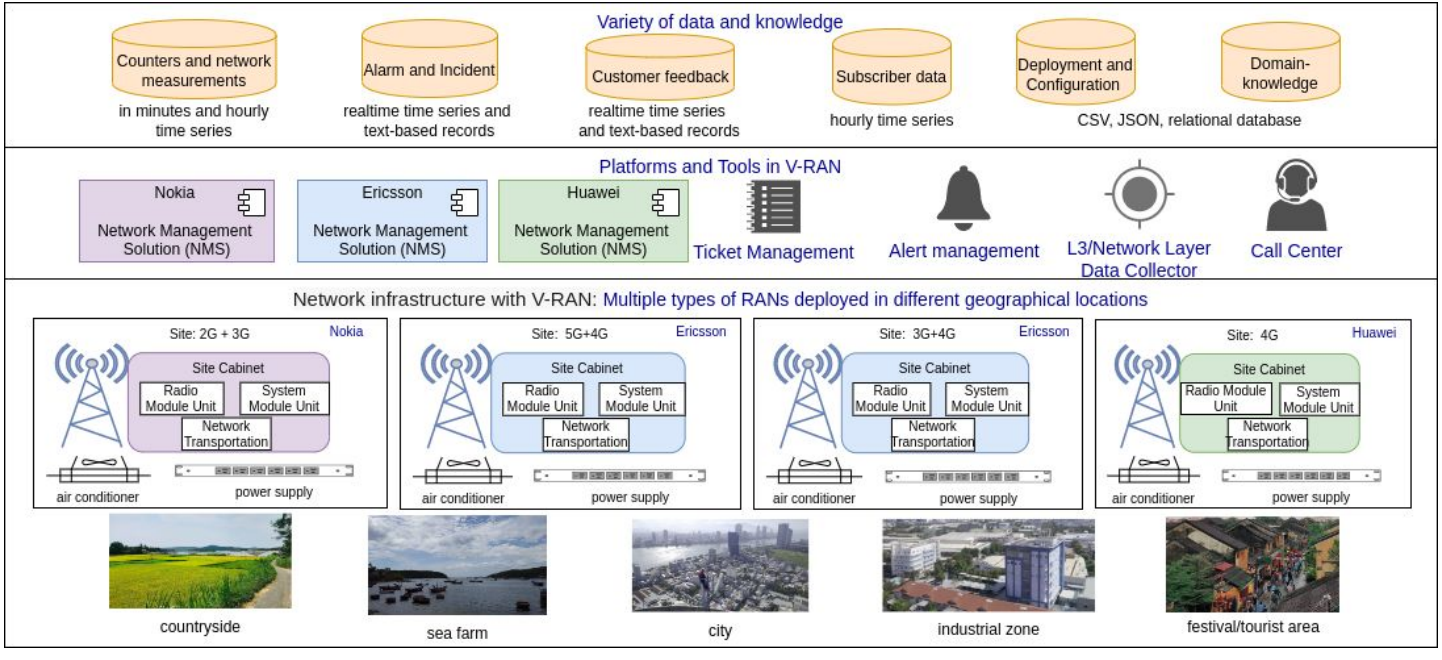
V-RAN - A Variety of Radio Access Networks

Diverse types of data & knowledge

Multiple existing isolated systems

Multiple vendors & technologies

Different business & operation contexts



Multiple V-RAN ***managed separately*** due to the organization structure

Holistic infrastructure needed, but many challenging issues

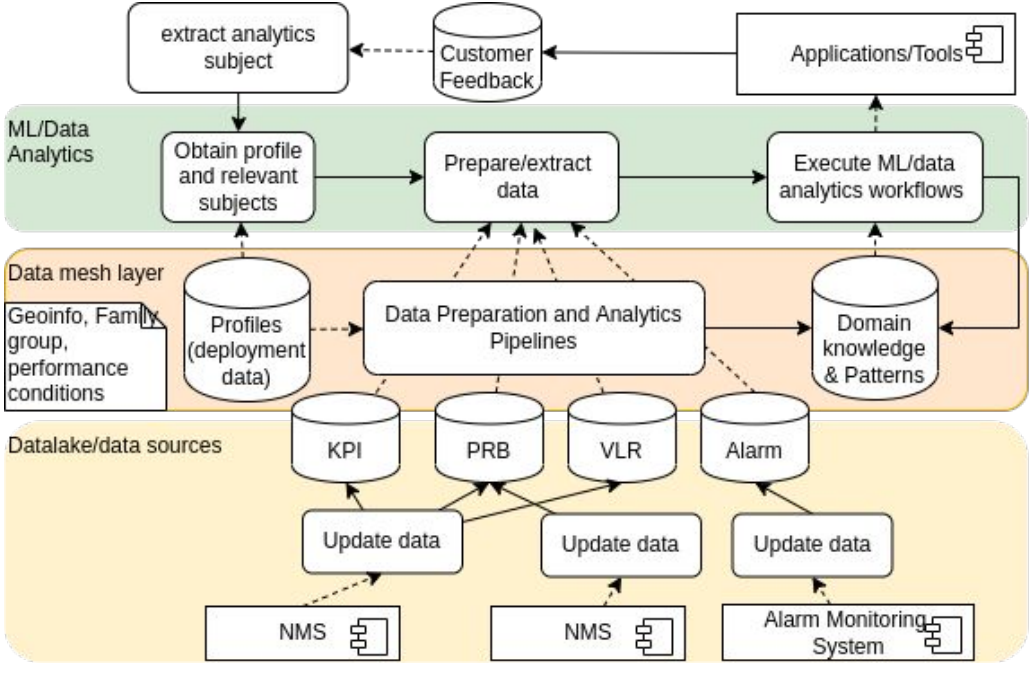
- **Many existing systems cannot be changed**
  - *Complex, but many are not designed for long-term big data analytics/ML*
- **Issues due to specific organization conditions**
  - *Engineers as ML dev/users: ML/data analytics are only for operations*
  - *Engineers lack ML skills but have extensive domain knowledge*
  - *Limited computing resources for ML but serving a large number of subscribers (e.g., ~4 millions of subscribers)*
- **Consequence**
  - *Under different contexts, the infrastructure must be designed suitable for organization operations*

Customized (light) data lake/mesh for quality assurance & engineer profiles

Holistic ML analytics for understanding diverse analytics subjects from multiple sources

Mesh layer is for holistic analytics, centered on analytics products suitable for different profiles of engineers

Customized quality-aware data lake designed only for data used in identified ML analytics

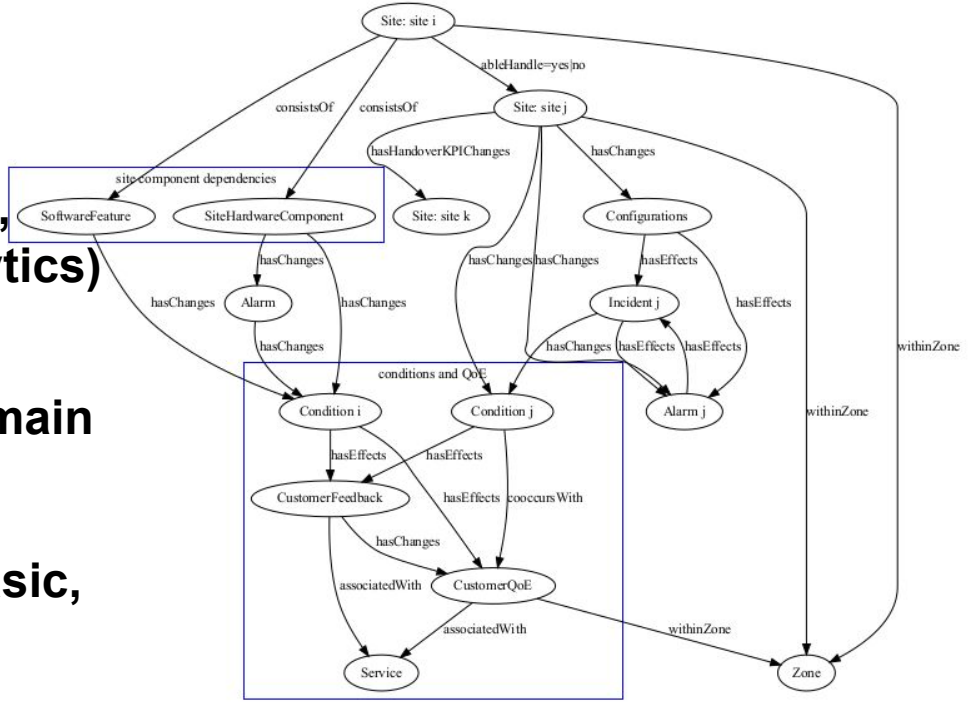


Enable diverse types of analytics subjects and dependencies

Analytics subjects associated with measurements (from monitoring), conditions (from domain knowledge), and statuses (from ML/big data analytics)

Holistic entities and relationships captured from data analytics and domain knowledge

A model of composable analytics: basic, light-weighted analytics for different types of subjects, before combining them→ easy to understand for engineers



Engineers need support for selecting suitable ML solutions

domain problems

datasets

domain knowledge

business & operation contexts

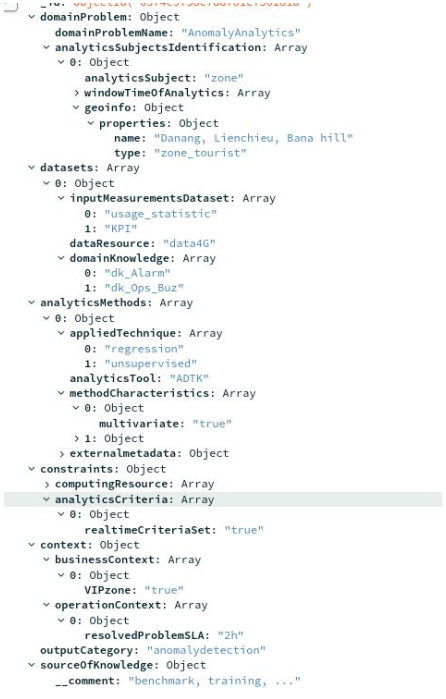
Enabling what, when, where and how for using ML analytics in operations

ML algorithms/ frameworks & evaluations

Carried out by ML engineer/data scientists to seek suitable methods

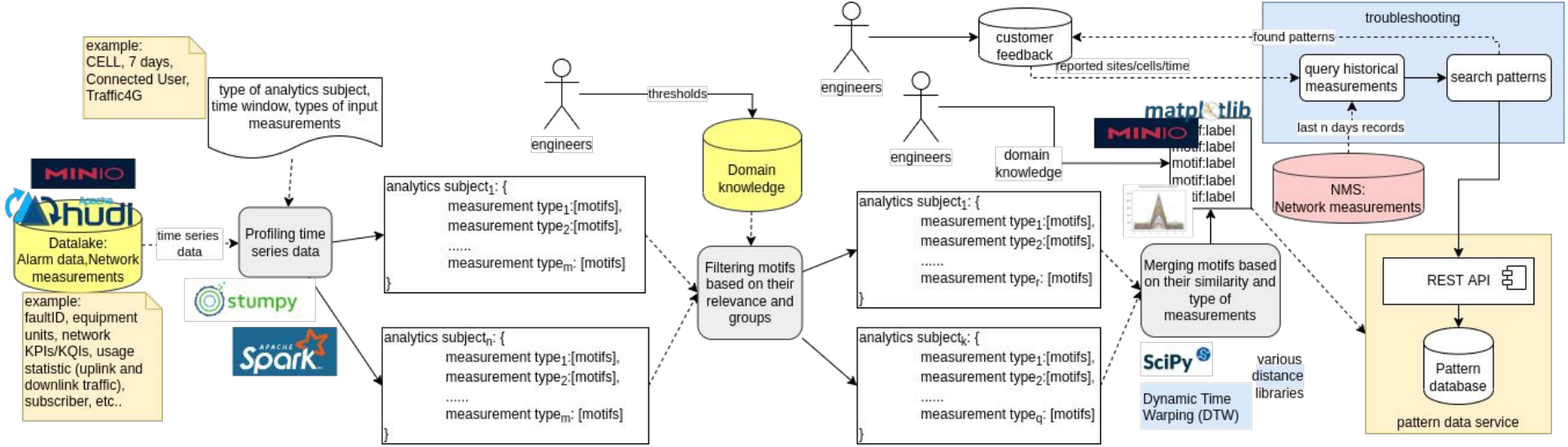
Identified, detected & provided by domain experts, engineers and big data analytics

Knowledge base for operations



Example: supporting troubleshooting of QoE

- **Continuously finding and updating quality patterns based on network traffic (ML) for near real-time feedback resolution (op)**



Conclusions

- **HAIVAN - a holistic ML framework**
  - supporting engineers in operating a variety of RANs, designed for constrained teams and infrastructures
  - focusing on data products, domain knowledge, models of analytics subjects and their dependencies, machine learning management for telcos engineers
- **Future work**
  - ML algorithms and couplings of ML analytics for analytics subjects

Thank you for your attention!

scan the paper